

Diversity in food technology

A scientific review, farm-scale trials and extensive public consultations on genetically modified crops should pave the way for greater benefits and choice for consumers — provided that the organic movement abandons self-damaging dogmas.

Last week in England, the Lake District National Park Authority, like other British regions before it, declared itself a GM-free zone. This came close on the heels of a meeting between Margaret Beckett, the British secretary of state for the environment, food and rural affairs, and heads of major retail chains. She was left in little doubt of the retailers' resistance to stocking genetically modified (GM) foods on their shelves, given customers' antipathy.

Ironically, these events coincided with the publication of a rather more positive scientific assessment of GM crops (see *Nature* **424**, 358; 2003). The review emphasized that, provided appropriate testing and regulation are in place for consideration on a case-by-case basis, GM crops hold out significant promise and leave little grounds for fear.

The next steps in the great British GM saga, which is being watched closely by many other countries, will be the publication of results of the farm-scale evaluation of oilseed rape, beet and maize, and the publication of the results of a major public consultation, both due in September. A final scientific review will then be produced for ministers. As the recently published review emphasizes, information from farm-scale evaluations is important in answering key questions about the effects of agricultural processes on wildlife.

The public debate warrants close scrutiny. The processes of consultations (some 40,000 responses) and public meetings (nearly 500, in all) are complete. But only now has the scientific review addressed an agenda of concerns set by initial public consultations. The succinct information provided on the website of the public debate and at meetings does not do justice to the messages now available from the science review. Although much public concern is focused on issues of ownership and equity, the late timing of the science review

limits the value of the public consultation on science-related issues.

More worryingly, open meetings in the public debate have been subjected to campaigning tactics by anti-GM lobbyists, leading to complaints from other members of the public that discussions have been compromised. So particular attention should be given to the independent evaluation of the consultation process.

The review left little doubt that the coexistence of GM and organic farming (assuming that approval for GM use is granted) will prove difficult to maintain. But the problem is an artificial one, based in essence on an ultimately arbitrary and self-defeating definition of 'contamination' by the organic movement.

Consider, for example, late blight in potatoes, a major problem for both conventional and organic farmers. Organic farmers contain the problem by applying copper sulphate-based preparations, which can harm the soil. Attempts to breed potatoes that are more resistant to the pathogen, *Phytophthora infestans*, have consistently failed to yield a marketable product. The best solution probably lies several years down the road in the next generation of GM crops.

British organic farmers — or at least the Soil Association, their campaigning organization — will resist seemingly to their dying breath the idea that inserting genes using molecular biology could be as ethical as the often less reliable but nevertheless technological approaches of conventional organic plant breeding and management. One can but hope that the messages from science will continue to be reassuring about the impacts of GM crops, and that they will combine with organic farmers' self-interest to demolish such phoney bastions, and allow both approaches to agriculture to prosper, in the ultimate interests of consumer benefits and choice. ■

Dangers of privatization

The Bush administration's drive to contract out services is a threat to science.

At major US archaeological centres, the buzz at this time of year is typically about some exciting discovery of America's past uncovered on a summer dig. But visitors this year to the Midwest Archaeological Centre in Lincoln, Nebraska, and to the Southeast Archaeological Centre in Tallahassee, Florida, will encounter gloom. These institutions, which for decades have provided archaeological analysis for the US National Park Service and other agencies, may be entering their final days.

The two centres have found themselves among the targets of a plan by the President George W. Bush's administration to privatize as many federal jobs as possible before the November 2004 election (see page 478). The centres serve as premier resources for federal land management, but their advocates fear that they are being privatized precisely because of their wealth of expertise: their studies may delay or halt mining, logging or road-building.

Republicans deny that there is an ideological game afoot, only moves towards more efficiency. But recent experiences invite scepticism. During the process of analysing which jobs might be privatized, a consultant told government administrators that the archaeologists

shouldn't be considered for privatization, as the centres' annual budgets are largely based on competitively secured projects. In other words, this is already lean science. But the consultant was told to keep his head down and avoid talking to congressional offices. A Republican Congressman from Nebraska described this appropriately as "a bean-counter doing something senseless".

An official at the Office of Management and Budget (OMB) told *Nature* that they have to consider archaeologists to be the same as laundry workers — not a sentiment likely to inspire confidence among *Nature's* readership. The OMB, after all, is the driving force behind the administration's privatization plans. Alarm bells should be sounding: the administration's zeal does not bode well for other agencies, with many scientists facing various degrees of privatization, including the National Institutes of Health and the Environmental Protection Agency.

Last week, administration officials backed away from some of their harsher privatization methods in the face of congressional opposition. Congress must continue to expose the Bush administration's privatization plans to tough scrutiny. ■





Into the maize
Genome map has
geneticists wondering
where to go next
p476



Pipe dream?
Russian scientists
fight for freedom
from spying charges
p477



Field studies
Can earthquakes be
predicted from
magnetic changes?
p478



Deep thoughts
Conservationists
uneasy over whale
population study
p479

Expert panel retreats from major restructuring in blueprint for NIH

Erika Check, Washington

The US National Institutes of Health (NIH) should be reorganized so that it can support high-risk, pathfinding research more effectively, says an eminent group of scientists and public-policy experts in what is likely to be an influential report.

The study was commissioned by Congress and carried out by the Institute of Medicine, part of the US National Academies (see *Nature* **418**, 572; 2002). It says that the director of the NIH should be given a formal role in the process of changing the number of institutes and centres — a switch from the present arrangement in which Congress takes the lead.

The authors also recommend that Congress should transfer the power to hire and fire institute directors from the secretary of the Department of Health and Human Services to the NIH director. In addition, they propose that Congress should give the director an annual fund of between \$100 million and \$1 billion for a programme to support “high-risk, innovative projects”.

But the authors, who released their proposals on 29 July, stopped short of endorsing the sweeping changes that some observers



Overall control: the US National Institutes of Health may be given greater autonomy.

had anticipated, such as the regrouping of the NIH's 27 institutes and centres into clusters based around body systems, such as the brain.

Some of the report's proposals are likely to prove controversial. The 21-member

panel, chaired by Harold Shapiro, former president of Princeton University, recommends merging the National Human Genome Research Institute, which has finished its major research goal, with the

Strategy for climate research gets cool response

Geoff Brumfiel, Washington

A much-criticized plan for guiding US research into climate change has been issued in its final form by the Bush administration. Many climate scientists say that it has been improved, but that a fundamental flaw remains — the plan lacks the budget and mechanisms to ensure that its results influence policy-making.

The strategy, unveiled on 24 July, is designed to coordinate research in important aspects of climate, such as sources and sinks in the carbon cycle and human influence on climate. Since the draft form was issued last November, it has been disparaged by many scientists, who say it ignores studies showing that climate change could damage the US

economy and environment (see *Nature* **420**, 110; 2002) and lacks a consistent framework to guide research.

Many changes have been made, and the report is now almost twice its original length, says James Mahoney, deputy administrator of the National Oceanic and Atmospheric Administration in Washington DC, who coordinated the report. Additions include five goals designed to guide research, including a focus on ecosystem response to climate change and past climate variability. The plan also calls for 20 reports over the next four years to provide guidance for politicians. “This is an intellectually sound document,” says Mahoney.

Critics agree that the plan is more

cohesive, but fear that the report will have little influence on politicians. “It's not just a question of science any more,” says James White, who chairs the Environmental Studies Program at the University of Colorado in Boulder. White says that the community needs to be talking to policy-makers about lessening climate change, not just issuing white papers.

Ultimately, budgets are the most pressing issue, adds Benjamin Preston, a senior scientist at the Pew Center on Global Climate Change in Arlington, Virginia. US funding for climate science is currently stagnating at about \$1.7 billion annually. Without fresh funds, much of the research in the plan will be impossible, he says. ■

► National Institute of General Medical Sciences. It also suggests re-evaluating the special status accorded to the National Cancer Institute, which has an unusual degree of independence from the NIH director.

Many researchers, including former NIH director Harold Varmus, have been calling for greater consolidation, claiming that the current arrangement of the NIH makes it inflexible and causes disparities in research funding. But changes to the status of individual institutes are likely to be opposed by lobbying groups and research organizations linked to the areas of science involved. The report's authors say such difficulties mean that further mergers would be politically impractical.

"Our discussions, correspondence and meetings made it quite clear that there would be very little agreement among these communities on what the right way to organize NIH is," the report's authors write, "and there would probably be dozens of conflicting ideas in play and few clear avenues for narrowing these down."

The authors also address the ongoing effort by the Department of Health and Human Services to centralize or outsource various NIH functions, such as some aspects of grant review. This move won the authors praise from the community for attempting to defend the NIH against interference from Congress and the Bush administration. David Baltimore, president of the California Institute of Technology, describes such behaviour as "gutsy". "They're taking on two big interests here — the administration and the Congress — and I think that's a very good thing," he says.

The question now is what will become of the report. A 1984 Institute of Medicine study, which included similar proposals, was not implemented by Congress. But politicians are currently more interested in the NIH than they were then. The agency is being investigated by Congress over payments to its researchers, for example. Last month, two members of Congress told NIH director Elias Zerhouni that they were beginning an investigation of payments for lectures made to NIH executives by large centres that received NIH money.

On 10 July, the investigation was expanded after an NIH programme administrator told reporters that even though he was removed from his role in 1995, he has since been paid an annual salary of \$100,000 while doing almost no work for the agency. This, together with the fact that Congress requested the report, makes it less likely that the study will be ignored, observers say. ■



Corny question: as the maize genome map nears completion, should a full genome sequence be next?

Maize map sees geneticists split over choice of direction

Carina Dennis

Plant scientists are entering the home straight in their bid to map the genome of maize (corn, *Zea mays*). Weighing in at 2.5 billion base pairs, the genome is about the same size as the human version, and will be the biggest plant genome yet mapped. The data are already accelerating the discovery of useful traits, but could also provide a springboard for a more ambitious effort: the detailed sequencing of the entire maize genome.

Current work is focusing on two maps. A high-resolution genetic map, analogous to a series of signposts showing the relative positions of different genes, was made available last November. A physical map, composed of overlapping cloned gene fragments to give the distances between the genes, is nearly complete and is being combined with the genetic version to create a scaled map of gene positions.

"We have assembled about 95% of the physical map and about half of it has been anchored to the genetic map," says Ed Coe, a geneticist with the US Department of Agriculture and a director of the Maize Mapping Project. Funded by a five-year, US\$11-million grant from the National Science Foundation (NSF), the project involves scientists at the University

of Missouri-Columbia, where Coe is based, and the universities of Arizona and Georgia.

Project scientists predict that the integrated map will be complete in September, but it has already helped researchers. "The map has had an impact on me," says Vicki Chandler, who studies gene control at the University of Arizona. "I now know exactly what the physical distance is between my genetic markers and where my gene is," she says.

For many crop scientists, the sequence of the maize genome is the next logical step. Others disagree, pointing out that the rice genome, which is already sequenced, is similar to that of maize. They argue that the rice sequence, together with the integrated maize map, will provide a good guide for geneticists.

But certain aspects of maize biology, such as the unusual vigour of some hybrid strains, need the genome sequence to be fully understood, some researchers contend. Gene order and the number of genes vary greatly between different strains, and this may contribute to hybrid vigour. "We can't use the rice genome as a reference sequence for understanding the diversity of maize," says Joachim Messing, a maize geneticist at Rutgers University in Piscataway, New Jersey.

Some experts think it would be impractical to sequence maize and other large, repetitive cereal genomes to the same standard as rice. Last September, to address these concerns, the NSF provided \$10 million for a programme to test a gene-enriching strategy for sequencing the maize genome. The project aims to develop a method for filtering out gene-poor areas of the genome, and to uncover ways in which the integrated map can be used to anchor data from sequencing of the gene-rich areas.

It remains to be seen what the private sector will contribute to the maize effort. Several firms, such as DuPont, based in Wilmington, Delaware, have sequenced parts of the maize genome. "At least two of the major players are interested in partnering with us," says Coe. ■



Russian scientists face long fight for justice

Byron MacWilliams, Moscow

Valentin Danilov is smoking a pipe and drinking black tea from mug that says, in Russian, "Don't touch me". On the mug, a black-and-white drawing of a thumb, pressing downwards, is surrounded by a red circle with a line through it.

In Danilov's case, such pleas are too late. A physicist at Krasnoyarsk State Technical University in Siberia, he was held in custody for 19 months after his arrest in 2001 for allegedly supplying a classified satellite-technology device to a Chinese company. Last week, his case continued its slow progress through the Russian legal system when the Supreme Court ruled that prosecutors should be given no more time to collect evidence, and that a lower court should reach a decision on the charges.

But no deadline has been set for that decision. The cases of several other scientists also remain unresolved. All date from between 1997 and 2002, when the FSB — the main successor to the Soviet KGB — accused around a dozen people, mainly scientists, of spying. With pressure from the Russian Academy of Sciences appearing to have little effect, human-rights campaigners are trying to increase awareness of what they regard as the persecution of researchers. "It has become fashionable to arrest scientists," Lyudmila Alexeyeva, head of the Moscow Helsinki Group, a human-rights organization, told a press conference in Moscow last week.

Danilov, like the other scientists under threat, has the support of his peers. His device is used to test the effect of electromagnetic activity on satellites, but the FSB believed that



From Russia with love? Valentin Danilov is charged with passing technology to China.

it could help China to develop spacecraft that would threaten Russia's security, although the agency did not say how. But in a letter to a Krasnoyarsk court, sent shortly after Danilov's arrest, 20 Russian scientists pointed out that details of the technology have already been published in scientific journals.

Other scientists have faced similar accusations. Vladimir Shchurov, who heads an acoustics laboratory at the Pacific Oceanographic

Institute in Vladivostok, was detained by customs officials in 1999 after attempting to deliver an underwater acoustic device to Harbin Engineering University in China. The FSB alleges that the device can be used for military purposes, such as tracking Russian submarines. The academy has written to the court handling the case to affirm that the technology was unclassified. Shchurov's trial is scheduled to begin this week.

Alexeyeva claims that Russian scientists have been an easy target because they collaborate with foreign companies and academic organizations. This, she argues, allows the FSB to accuse them of spying, in the hope that the publicity will deter any researchers who are considering selling classified information from doing so. The FSB had not responded to questions faxed to its offices in Moscow as *Nature* went to press.

None of the scientists has yet been convicted of spying, however, and some have had their cases dismissed. Last week, charges were dropped against Vladimir Soifer, an ecologist at the Pacific Oceanographic Institute accused of handling classified documents about radioactive pollution from an accident involving a nuclear submarine in Chazhma Bay, near Vladivostok, in 1985. *Nature* has obtained a copy of a letter sent to Soifer on 18 July, in which V. A. Zhilyaev, a lieutenant general in the FSB, apologizes for the accusations.

Despite the pending case, Danilov says the stigma of being charged has not affected his personal or professional life. Some colleagues even provided financial assistance to his wife while he was imprisoned. "I walk with my head held high," he says. ■

Anger grows over plan to uproot Indian crop institute

K. S. Jayaraman, New Delhi

A row is brewing over plans to relocate the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), currently based in the Indian state of Andhra Pradesh. Plans to move the centre — one of a global network of 16 Future Harvest centres dedicated to food security and environmental issues — have angered staff and scientists across India, who fear that this will disrupt its research projects.

The proposal came in a report released earlier this month from an independent five-member panel set up by the Washington-based Consultative Group on International Agricultural Research (CGIAR), which runs the Future Harvest centres. The panel said that ICRISAT's outstations in Kenya, Niger and Zimbabwe need strengthening, and that the best way to



Pick of the crop: Indian scientists have worked to improve pearl millet, a staple in the tropics.

do this is to relocate the institute to Africa.

ICRISAT, which carries out research on crops that are important to the arid regions of Africa and India, such as millet and groundnut, has claimed some notable successes, such as trebling chickpea yields by

introducing new varieties. But the panel concluded that its research would have more impact if it were based in Africa.

Budget cuts forced the loss of some 200 positions at ICRISAT last year, but it could be spared further job losses, as the panel concluded that the institute's programme to identify beneficial genes and improve crops through conventional breeding and genetic modification should still be based in India.

The institute will respond to the relocation plan when its board meets in September. "This recommendation is unfortunate," says Raj Paroda, director of the CGIAR programme for central Asia and the Caucasus. He adds that it would adversely affect the growth and development of the institute, as India has stronger national research organizations than Africa. ■

♦ www.icrisat.org

Satellites aim to shake up quake predictions

Tony Reichhardt, Washington

The controversial idea that earthquakes can be predicted by monitoring tiny fluctuations in Earth's magnetic field is to be tested by two new satellites.

Although many seismologists see little merit in the idea, NASA and the US Air Force are together contributing about \$1 million to provide data analysis and ground instrumentation to support experiments with the first satellite, the privately funded QuakeSat. Built by QuakeFinder of Palo Alto, California, the craft is now returning data from orbit after its 30 June launch. A second more expensive and ambitious satellite, funded by the CNES, France's national space agency, will follow next April.

Both craft are designed to search for small variations in Earth's magnetic field, which some scientists believe precede earthquakes. According to their theories, extremely low-frequency magnetic fluctuations are produced by compression of crystalline rocks or the movement of water in fault zones before a quake. Such fluctuations could potentially be detected from an orbiting satellite.

The idea of using magnetic activity to predict earthquakes is not entirely without precedent. Antony Fraser-Smith, a geophysicist at Stanford University in California who is among those who will receive NASA funds to work on QuakeSat data, reported elevated ground-based magnetic signals several days before the 1989 Loma Prieta earthquake, which damaged San Francisco and other cities in California (A. C. Fraser-Smith *et al. Geophys. Res. Lett.* **17**, 1465–1468; 1990). Most Earth scientists consider these results to be interesting but inconclusive.

There is also tentative evidence that such fluctuations could be detected from space.



Satellite studies will look at the idea that magnetic signals preceded the 1989 Loma Prieta quake.

Russia's Cosmos-1809, for example, managed to record signals from a 1989 earthquake in Armenia. Although some other satellites have turned up no such evidence, the subject intrigues geophysicists.

QuakeSat, built in collaboration with Stanford's Space Systems Development Laboratory, will now test these ideas. The satellite, which will collect data on magnetic-field strength from orbit for about six months, will have greater coverage than ground-based magnetometers — which are QuakeFinder's main product.

The company's president, Marijean Seelbach, admits that "the jury's still out" on whether the device can predict earthquakes. Stephen Park, a geophysicist at the University of California, Riverside, is typical of those

who doubt that it can. Magnetic fields during the main shock of an earthquake generally measure billionths of a tesla at the fault line, he points out. Precursor fluctuations are likely to be lower, and would lose energy as they travel to the satellite. Seelbach agrees that measuring such tiny signals is a daunting task, but says that QuakeSat will be sensitive to trillionths of a tesla.

NASA is also "agnostic" about whether the satellites will prove useful, says John LaBrecque, manager of the agency's Solid Earth and Natural Hazards research programme in Washington. He admits that NASA would have had difficulty being the sole funder for such a controversial project, but says he feels the agency is justified in making a contribution to a private scheme. ■

US researchers fear job losses from privatization drive

Rex Dalton, San Diego

US scientists have reacted anxiously to a government plan to contract out federal scientific projects, citing fears that the scheme could damage research. The plan met stiff opposition in Congress earlier this month, but the Bush administration wants to implement it before the November 2004 presidential election.

The administration's plan involves contracting out about 425,000 federal jobs, which may include hundreds of researchers. Under the scheme, private firms with scientific staff might perform research that is now conducted by government scientists. Administration officials say that privatizing jobs will save money and increase efficiency.

The scheme hit problems on 17 July, when the House of Representatives voted to block plans to privatize two National Park Service regional archaeology centres, involving 100 staff. Researchers at the Midwest Archeological Center in Lincoln, Nebraska, and the Southeast Archeological Center in Tallahassee, Florida, have been carrying out cultural studies on public and private lands for more than 30 years.

But further disputes are expected in the coming weeks, as congressional hearings for the 2004 financial-year budget discuss privatization in other agencies, such as the National Institutes of Health (NIH) and the Environmental Protection Agency (EPA). One union official at the EPA says he fears

that 150 jobs in his 1,000-member chapter could be privatized. The NIH is already studying how to outsource almost 1,000 jobs in the 2004 fiscal year.

Critics of the plan say that staff whose positions are privatized could lose their jobs if they are not re-employed by the private firm that takes over. Others add that private companies could be more willing to bow to political pressure over controversial research. "When you replace government scientists with private ones, the latter are more likely to pull their punches, and not raise or address sensitive issues," says Jeff Ruch of Public Employees for Environmental Responsibility, a Washington-based pressure group that promotes government accountability. ■



Making a splash: an estimate of historic humpback-whale populations has divided biologists.

Whale genetics study leaves conservationists all at sea

Tom Clarke and Jonathan Knight

Conservationists fear that a study on whale populations, which seems to back the need for more restrictions on whaling, could ultimately undermine their cause.

The analysis, published on 25 July, suggests that the population of humpback and fin whales before widespread whaling began in the nineteenth century was about 12 times larger than previously thought (J. Roman and S. R. Palumbi *Science* **301**, 508–510; 2003). This could have important implications, as the International Whaling Commission (IWC) says that commercial whaling can resume when populations approach their pre-whaling size. But critics say the analysis is flawed and should not be used to set policy.

Estimates of past population levels are usually based on whaler's logbooks and population growth rates. Palumbi looked instead at the genetic variation in current whale populations by comparing the sequence of a small region of DNA collected from several hundred animals. The greater the degree of variation, the larger the historic population.

The analysis suggests that the North Atlantic once supported 240,000 humpbacks, far more than the IWC's pre-whaling estimate of 20,000. There are about 10,000 humpbacks in the North Atlantic today.

"The numbers strongly caution against the return to whaling for humpback and fin whales," says Stephen Palumbi, a population biologist at Stanford University in California, and one of the study's authors.

Yet some whale experts say that Palumbi's

conclusion is a gross overestimate. "It's completely out of the realm of reality," says Phillip Clapham, a whale biologist at the Northeast Fisheries Science Center in Woods Hole, Massachusetts, who opposes the resumption of commercial whaling. "To suggest that these figures be used as a target for the managed recovery of whale populations is inappropriate," he says.

One problem is that the estimate of how frequently genetic mutations occur has a big influence on the analysis, says geneticist Per Palsbøll of the University of California, Berkeley. A higher but also plausible mutation rate would have shown that the pre-whaling humpback population was only 2,400. Interbreeding between southern and northern populations could also have increased the genetic diversity, making historic populations seem larger than they were, he says.

Palumbi says that the mutation rate he used is consistent with the evolutionary history of whales and related species, adding that interbreeding would not significantly alter his results. He notes that other techniques have not been peer-reviewed or subject to the scrutiny that his paper is undergoing.

If pro-whaling interests seize on the controversy, says Richard Mott, vice-president for international policy at the environmental organization WWF in Washington DC, it will be important to point out that debate is part of good science. "This shows that on the conservation side the standards for debate and peer review are very rigorous, which is not the case in the pro-whaling countries," he claims. ■

US bombshell hits revamp of Russian weapons centres

Geoff Brumfiel, Washington

A US-funded scheme to turn the secret cities that developed Russia's nuclear arsenal into commercial centres could come to end this September.

Officials at the US Department of Energy (DOE), which runs the scheme, say that the Nuclear Cities Initiative must close unless Russia accepts liability for American workers and companies on the project. But arm-controls experts charge that this is a convenient excuse used to kill a programme that is unpopular with the Bush administration.

The scheme, which was born in 1998 and currently consumes around US\$15 million in funds annually, has been hampered by problems. Some Russian weapons centres have been reluctant to grant access to their sites, and a 1999 report by the US General Accounting Office said that too little money was reaching Soviet scientists.

Despite this, the initiative has support within the arms-control community, as it is the only US programme aimed at converting weapons facilities to civilian uses. The scheme has had some early successes, such as the shutdown of a nuclear warhead production plant in the formerly closed Russian city of Sarov.

But on 22 July, US energy secretary Spencer Abraham told Alexander Rumyantsev, Russia's atomic-energy minister, that he will not renew the agreement when it expires in September. In a statement, Abraham said the scheme could not continue unless Russia agreed to accept legal liability for US companies and scientists working in the cities, including protection from prosecution in the event of an accident. The Russian parliament is currently in recess and is unlikely to to address the issue before the September deadline has passed.

Supporters of the programme question the administration's motives. "The liability issue is a red herring," says Rose Gottemoeller of the Carnegie Endowment for International Peace in Washington, who led the DOE's non-proliferation bureau under the Clinton administration. Gottemoeller points out that the Bush administration tried to cut the programme in 2001, but that funding was restored after the terrorist attacks of 11 September 2001.

US energy undersecretary Robert Card says that the DOE will back currently funded projects to completion. ■

♦ www.nn.doe.gov/nci

Ecologists count cost as wildfire ravages research station

San Diego A climate and ecosystem research facility in southern California has been torched by a wildfire raging through mountains north of San Diego. The Sky Oaks Field Station, in an area of chaparral shrubs, lost equipment and samples worth about \$1 million on 17 July, according to early estimates.

Research teams from San Diego State University (SDSU) and California State University, Los Angeles (CSULA) had been using robotic devices for seven years to test atmospheric gases in relation to plant respiration and photosynthesis at the site.

SDSU's Global Change Research Group, led by ecological physiologist Walter Oechel, will see how much research can be salvaged. Plant physiologist John Gamon of CSULA's Center for Environmental Analysis says that tramways used to move the robots across the site were wiped out. But he now hopes to add a study of the fire's impact on the chaparral ecosystem to his group's project.

Court rules against NASA over Hawaiian telescope

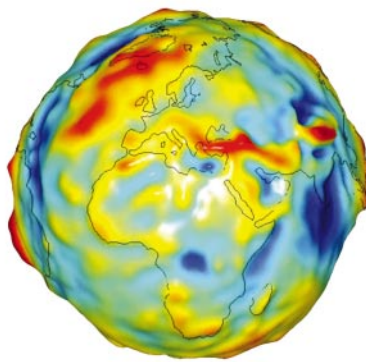
Washington More than a year after a Hawaiian state agency sued to stop construction of NASA telescopes on the traditionally sacred Hawaiian mountain of Mauna Kea, NASA has received the decision it was dreading. The US District Court in Hawaii ordered NASA on 15 July to prepare a full environmental-impact study before building new 'outrigger' telescopes for the Keck Observatory.

The proposed group of four to six small telescopes would improve Keck's ability to see fine detail, which NASA would use in the search for planets around other stars.

Spokesman Don Savage says that the agency is still evaluating its options. NASA may even scrap the project, which has been delayed by wrangling between astronomers and Hawaiian natives (see *Nature* 417, 5; 2002). An environmental-impact study is expected to take months and to add greatly to the \$20 million that NASA has already spent developing hardware for the outriggers.

Gravity map gets to grips with planet's pull

Washington The most accurate map yet of Earth's gravitational field has been released — the first result of the Gravity Recovery and Climate Experiment (GRACE), run jointly by NASA and the DLR, Germany's aerospace research centre. Oceanographers hope to use the map to model deep oceanic currents and heat fluxes, and so better



Use the force: the raised, red areas in GRACE's gravity map show where Earth's pull is strongest.

understand their effects on global climate.

"It's an order of magnitude better than anything that's gone before," says Byron Tapley, the project's principal investigator and director of the Center for Space Research (CSR) at the University of Texas at Austin. The map, called GRACE Gravity Model 01, was based on a preliminary analysis of 111 days of mission data. It was published on the CSR website on 21 July, a few months ahead of the main phase of the mission.

The GRACE mission comprises two identical satellites, one 220 kilometres ahead of the other. The lead satellite encounters local variations in the gravity field first, and so temporarily pulls away from its partner (see *Nature* 416, 10–11; 2002). The map was created by following these minute changes in separation, which are measured to the nearest micrometre by onboard sensors.

Row over junk DNA patents comes home to Australia

Sydney Genetic Technologies (GTG), the Melbourne-based biotech company that has patents on the use of non-coding DNA, has granted a research licence to the University of Sydney. This is the second of its academic licences, and the first on its home turf.

The company has drawn criticism for requiring academics to buy licences for the use of 'junk DNA' sequences in genetic

analyses (see *Nature* 423, 105; 2003). Francis Collins, who heads the National Human Genome Research Institute in Bethesda, Maryland, spoke out against GTG's policy at the International Congress of Genetics in Melbourne last month.

GTG charged US\$1,000 each for the two licences sold so far. "Our intention is simply to regularize researchers' activities so they are no longer infringing our patent," says Mervyn Jacobson, chairman of GTG. "We separate research licences for academics from commercial licences for organizations selling products in the marketplace — we are not planning to profit off academic researchers."

The University of Utah in Salt Lake City was the first to buy a research licence, in May. Other institutions are also seeking licences, says Jacobson.

Mammal experts get networking opportunity

San Diego Data on 1.4 million specimens from mammal collections throughout the United States, Canada and Mexico are to be linked by a new computer network.

The Mammal Networked Information System will allow researchers to retrieve data remotely from any of the 17 participating institutions. The system should help users to monitor ecosystems, identify invasive species and track diseases that move from mammals to humans. Researchers studying a 1993 hantavirus outbreak in the US southwest, for example, used samples from mammal collections in the region to trace the strain to the deer mouse, a common North American rodent.

The project is led by the Museum of Vertebrate Zoology at the University of California, Berkeley, aided by a \$1.4-million grant from the National Science Foundation. Eleven of the mammal collections are now online, with the remainder to be included by early next year.

♦ <http://elib.cs.berkeley.edu/manis>

Endangered footage finds an online ark

London Recorded images and sounds of endangered species are often as threatened as the animals and plants themselves. Wildlife films, for example, are often lost when production houses close or technology changes.

But a major digital storage project called ARKive is now giving such records a permanent and safe home. ARKive aims to accumulate records on all 11,000 species listed by the World Conservation Union as at risk of extinction, and make them available online.

The project, based in Bristol, UK, is funded by Britain's Heritage Lottery Fund and by Hewlett-Packard Research Laboratories Europe.

Highlights already stored in ARKive include the



only film of the extinct Tasmanian tiger (pictured), and images of the coelacanth, a fish that has roamed the seas since the days of the dinosaurs.

♦ www.arkive.org

Setting the record straight

By analysing masses of data from fossils throughout the world, a group of palaeontologists hopes to address the big questions about the history of life on Earth. Quirin Schiermeier logs on to the Paleobiology Database.

We've never had it so good — or at least that has been the prevailing view among palaeobiologists who have tried to track the history of our planet's biodiversity. On the long road from the first stirrings of multicellular life to today's shimmering diversity, untold numbers of species have fallen by the wayside. From time to time, legions of creatures have perished together in mysterious mass extinctions. But if you examine the fossil record, the evolution of new species seems generally to have had the upper hand over extinction. Like stock indices in a bull market, plots showing the diversity of life over geological time reveal a rising trend, despite occasional setbacks.

But how can we be sure that this isn't a sampling artefact? Even high-school biology students are taught that the fossil record is far from complete. Given that younger rocks are more likely to be exposed at the surface, it is possible that the apparent rise in biodiversity merely represents the greater scrutiny that has been applied to these strata. Palaeontologists have even coined a term for this source of bias: 'the pull of the recent'. Add in the confusion caused by the varied names used to describe the same organisms, and some researchers argue that attempting to assess the history of Earth's biodiversity is a fool's quest.

John Alroy, a palaeontologist at the University of California, Santa Barbara (UCSB), begs to differ. He is one of the founders of the Paleobiology Database, a project set up in 2000 with financial support from the US National Science Foundation. This freely accessible database, hosted by UCSB's National Center for Ecological Analysis and Synthesis, already holds information on more than 30,000 different fossil collections, and is still growing. Through sheer weight of numbers, and by applying various statistical tricks to account for sampling biases, Alroy and his colleagues hope to determine whether the Earth's bio-

diversity really has been on the rise — and to answer some other tricky questions.

The data are divided into individual collections, retrieved from specific locations and strata by particular palaeontologists. In addition to descriptions of specimens, the database includes information on the composition and age of the sediments in which they were found, and the fossils' state of preservation. "It is the multitude of easily retrievable information that makes it so useful," says Wolfgang Kiessling, a palaeontologist at the Museum of Natural History in Berlin, who is one of the 70 or so scientists authorized to enter information into the database. "It allows us to interpret the known fossil record in a more unbiased way, and we can now ask a whole host of new questions about how natural systems operate."

But some sceptics suspect that no amount of statistical sophistication will eliminate the uncertainty inherent in palaeontology. It may always be impossible to determine the degree to which the fossil record is 'known', they argue. And some fear that the Paleobiology Database will seduce unwary researchers into drawing erroneous conclusions. "No doubt palaeontology will benefit from more informed fossil data," says Andrew Smith, an invertebrate palaeontologist at London's Natural History Museum. "But you can easily be misled if you assume that all data are objective."

Multicellular life left little impression on the fossil record for around half a billion years. But at the onset of the Cambrian period, some 550 million years ago, oxygen and calcium had become sufficiently abundant in the oceans for the development of organisms with hard components. The result was the 'Cambrian explosion' of marine biodiversity. Our understanding of diversity trends since then owes a heavy debt to the work of John 'Jack' Sepkoski, a palaeontologist at the University of Chicago. In the 1970s, Sepkoski began to scour the palaeontological literature for information on the first and last appear-

ances of marine organisms, extrapolating the ups and downs of life in the oceans. Because individual species appear only fleetingly in the fossil record, and are often misidentified, he focused on genera — the trilobites *Paradoxides bohemicus* and *Paradoxides gracilis*, for instance, are species within the same genus, whereas *Asaphus cornutus* and *Crozonaspis struvei* belong to different genera.

Expanding knowledge

Sepkoski's work¹ indicated that diversity continued to expand from the Cambrian explosion until the end of the Ordovician period, some 440 million years ago. From then on, it remained on a more or less stable plateau until the Permian-Triassic mass extinction — the most severe experienced by our planet. But Sepkoski found an upward trend in marine biodiversity after the beginning of the Triassic period, 250 million years ago, with just one significant interruption: the extinction event that put paid to the dinosaurs at the end of the Cretaceous period, some 65 million years ago (see figure, opposite). "Sepkoski has given us a clue about the dramatic things that happened in the history of life," says Alroy. "But we need more comprehensive data, and better analytical tools, to quantify his findings."

This is what the Paleobiology Database aims to provide. Because of the information included about individual collections, it is possible to correct for sampling biases in ways that



John Alroy hopes the database will answer crucial fossil questions.



Rock compilation: snapshots of fossil diversity are being united to reveal the bigger picture.

and his colleagues argued, the increase in diversity over time does seem to be real.

Alroy and his colleagues believe that the database is the key to resolving this and other controversies surrounding the history of life on Earth, such as whether the great mass extinctions really were as dramatic as has been assumed. It should also give palaeo-ecologists a better idea of whether biodiversity is controlled by environmental parameters such as climate, volcanic activity and ocean chemistry — or whether, as a theory proposed two years ago⁴ suggests, it varies randomly.

Share and share alike

Getting the most out of the database may require a cultural change on the part of some palaeontologists, however. Further expansion of its scope will require researchers to make their collections available for analysis. The situation in New Zealand, where palaeontologists began three decades ago to compile and publish all fossil data in an openly searchable way, under an agreement between the Geological Society of New Zealand and the New Zealand Geological Survey, provides an ideal model. In a paper published just a few weeks ago, these data were used to study bias in measurements of mollusc diversity caused by variation in the total area of exposed rocks of different ages⁵. But in Europe, says Kiessling, some palaeontologists still jealously guard their own collections to maintain an advantage over their rivals.

The value of the Paleobiology Database will depend on the quality, as well as the quantity, of its information. Some experts fear that quality-control issues could cause misleading results, particularly in the hands of scientists who are not experts on the organisms that they are trying to analyse. Smith, for instance, is concerned about the potential for confusion due to problems with taxonomic nomenclature. "Names may disappear, but their last occurrence in the record does not necessarily mean extinction of a species, family or genus," he says. Despite such shortcomings, however, Smith intends to use the database and contribute to it. "But I would only work with taxa that I know," he adds.

Alroy and his colleagues are trying to address the problem that Smith has highlighted. At a meeting later this year, they plan to set up task force to resolve inconsistencies in the database. Once this group's work is done, enthusiasts claim, the database will be a powerful tool. "It will add a long-term perspective to many open questions," says Kiessling. ■

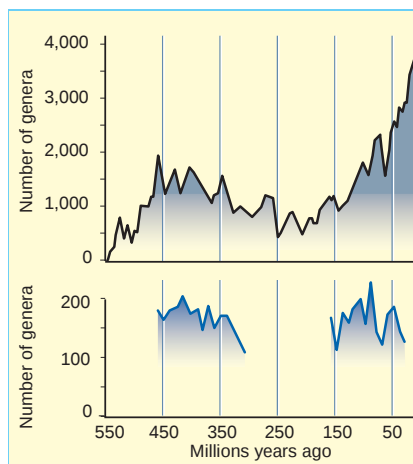
Quirin Schiermeier is *Nature's* German correspondent.

1. Sepkoski, J. J. Jr *Paleobiology* **7**, 36–53 (1981).
2. Alroy, J. et al. *Proc. Natl Acad. Sci. USA* **98**, 6261–6266 (2001).
3. Jablonski, D., Roy, K., Valentine, J. W., Price, R. M. & Anderson, P. S. *Science* **300**, 1133–1135 (2003).
4. Hubbell, S. P. *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton Univ. Press, Princeton, New Jersey, 2001).
5. Crampton, J. S. et al. *Science* **301**, 358–360 (2003).

♦ www.paleodb.org



SOURCE: REF. 2



On the level: the traditional view that biodiversity has increased over time (top) is challenged by an analysis that corrects for sampling bias.

Sepkoski, with his simple analysis of the first and last appearances of genera in the fossil record, was unable to do. For instance, within each interval being studied you can examine a fixed number of collections, in an attempt to account for variation in sampling intensity from rocks of different ages. Other corrections can be applied to account for bias in geographical coverage, and so on. Sepkoski himself realized the need to do this, and contributed to the Paleobiology Database until his death from heart failure, aged just 50, in 1999.

Initial analyses have already given some intriguing hints of discrepancies from Sepkoski's earlier conclusions. In the first major paper² to make use of the Paleobiology Database, Alroy and 24 colleagues — including the late Sepkoski — sampled the database's marine component, which at the time contained 8,591 collections, mainly from North America and Europe. They applied four different statistical methods to correct for variation in sampling intensity in rocks of different ages. Each gave roughly the same result, suggesting that marine biodiversity has not risen over the past 150 million years, and is at a similar level to that during the period between 450 million and 300 million years ago (see figure).

If this finding is correct, it means that Sepkoski's conclusion about rising diversity since the Triassic is a sampling artefact. "This is a very important and surprising result," says Kiessling. "It is the first time evidence has been found that there may be an upper threshold to biodiversity — a maximum holding capacity of the environment." The threshold theory is controversial, however, and the picture may yet change again, as researchers consider data on other taxonomic groups or from different regions.

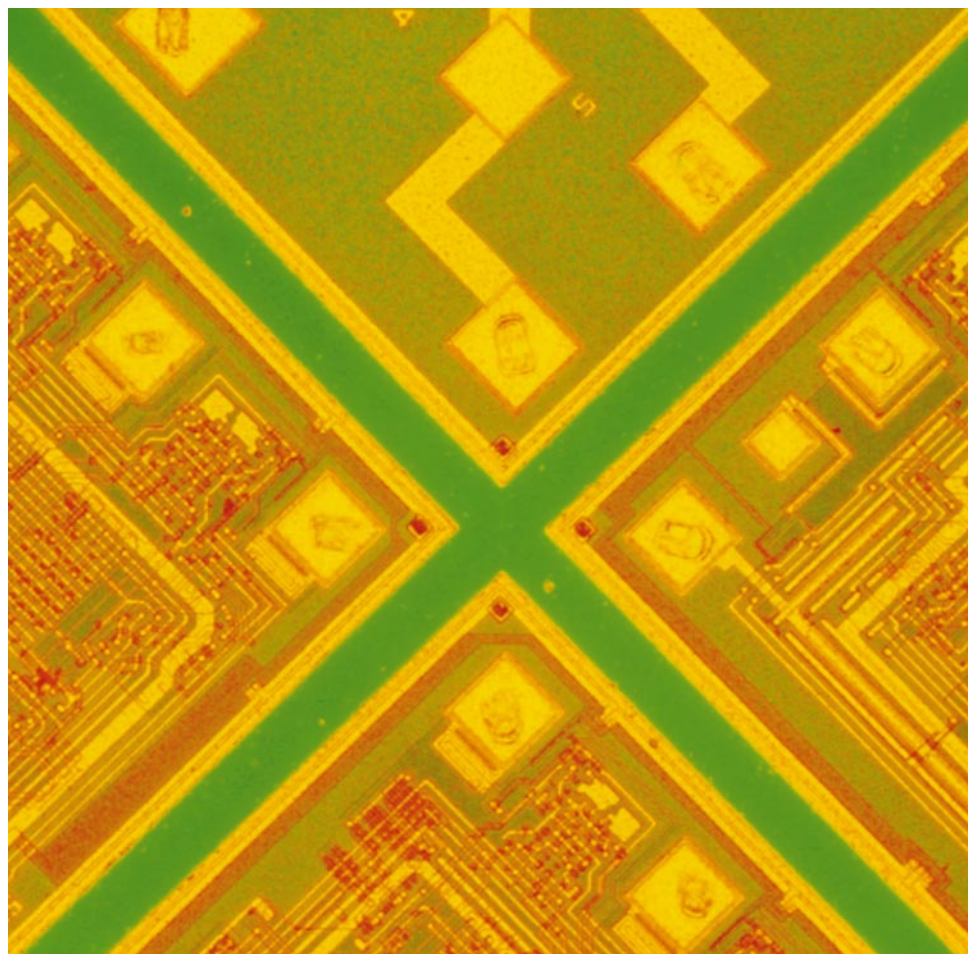
Indeed, a team led by palaeontologist David Jablonski of the University of Chicago has analysed the database's entries for bivalve molluscs, concluding that the pull of the recent has been overestimated in previous studies³. For this group of animals, at least, Jablonski

Quantum bits and silicon chips

Quantum computers offer a new kind of processing power. Silicon chips are easy to manufacture. Can the advantages of the two approaches be combined? Jenny Hogan reports.

For all their promise of fast and furious number-crunching, prototype quantum computers are bulky and impractical beasts. Some fill rooms but remain feeble calculators. Others require precision manufacturing that stretches current techniques to the limit. And most must be cooled to within a whisker of absolute zero before they can be used. They may have extraordinary potential, promising to crack complex codes and solve age-old mathematical puzzles, but the devices seem unlikely to slot neatly into desktop machines.

Not unless Marshall Stoneham is right. A materials scientist at University College London, Stoneham is using his experience in semiconductors to marry the worlds of silicon chips and quantum computing. He isn't the first to pursue the silicon approach, but the design he favours can, in theory, be manufactured with the tools available today. It should be powerful enough to do useful calculations and ought to operate at higher temperatures than rival devices — perhaps even at room temperature. Earlier this year, his team won £3.7 million (US\$6 million) from the UK government-funded research councils to put his idea into practice. Confident in his novel design¹, Stoneham taps his desktop computer



Thought processors: could today's silicon chips provide the basis for ultra-fast quantum computers?

and says: "If we succeed, we'll have a quantum processor on people's desks by 2010."

Like all quantum computers, Stoneham's design shares a common feature with conventional processors — both carry out calculations by manipulating bits of information, the ones and zeros of binary code. The simplest operation involves flipping the value of a bit, so that 1 becomes 0 or vice versa. More complex operations involve two bits. One bit may be flipped if the value of a second bit is 0, for example, but not if it is 1.

In principle, quantum devices work in the same way. But they also take advantage of the rules of quantum mechanics, which say that particles can exist in two states simultaneously. An electron that moves from one atom to a neighbouring atom, for example,

can actually be in both positions at once. If the electron is used to represent a bit, so that it is given the value of 0 when it is next to one atom, and 1 if it is next to the other, interesting possibilities emerge. Applying an external influence such as an electric field to the electron can flip the value of the quantum bit, or qubit. But because the electron is in two places at once, the operation is performed simultaneously on both states.

When qubits are linked together, this becomes a significant advantage. Two electrons, for example, can exist in four different states — 00, 01, 10 and 11 — depending on the relative positions of the particles. If the electrons interact with each other — a phenomenon called entanglement — then any operation carried out on one electron will

D. PARKER/SPL

P. SALOUTOS/CORBIS; T. CRADDOCK/SPL

simultaneously be carried out on the other. In effect, this means that one operation is carried out on four different states at the same time. Add more qubits and the number soars — a computer with just 30 qubits can perform more than a billion operations at once.

Not all computing tasks would benefit from such 'parallel' processing. A word-processing program, for example, probably wouldn't be enhanced by quantum hardware. But theory suggests that certain problems, such as factorizing large numbers — a crucial part of code-breaking — would yield to quantum techniques.

There are several ways to build a quantum computer, each of which stores and manipulates qubits in a different way. One method stores qubits in the energy states of an ion, and simple programs have already been run in this way. This January, for example, Rainer Blatt and his colleagues at the University of Innsbruck in Austria chilled a calcium ion to within a few millikelvins of absolute zero, and manipulated its energy states using laser pulses². One qubit was stored in the energy levels of the ion's motion, the other in the energy of its orbiting electrons. Blatt's team ran a simple algorithm that calculated whether a particular function always gives the same output, but researchers believe they will soon be able to combine several ions to perform more complex calculations.

In a spin

Others are using the spin of atomic nuclei — a quantum-mechanical property that aligns itself with or against an external magnetic field — to represent a 1 or 0. By adapting the nuclear magnetic resonance (NMR) technology that is used to image molecular structures, Jonathan Jones, a physicist at the University of Oxford, UK, created a liquid-based two-qubit device in 1998. The qubits were stored in the nuclear spins of hydrogen atoms in the biological molecule cytosine. The nuclei were constrained by strong magnetic fields, and radio waves were used to flip their spins³. The result of the calculation can be deduced from the spectrum of the interacting radio waves. NMR computers have since been extended to create the most advanced quantum computer so far — a seven-qubit device⁴.

But despite such progress, these systems have drawbacks. The equipment needed to manipulate ions fills a room, and the magnet-



Marshall Stoneham believes that he can make a viable desktop quantum computer by 2010.

ic fields used in NMR are generated by huge superconducting magnets that require cooling with liquid helium. The nuclear-spins approach also runs into trouble as the number of qubits increases, because the result signal gets lost in the noise from neighbouring molecules. In effect, the computer crashes if it uses eight or more qubits. Even with seven qubits "it begins to fall apart", says Jones.

But would nuclear spins be easier to control if they were embedded in a solid? Silicon is the ideal substance with which to test this idea. Decades of investment in conventional computers have generated a wealth of knowledge about how to fine-tune its properties, and theory suggests that silicon atoms should not interfere with the nuclei of whatever element is chosen to carry the qubits.

Up to five years ago such a system was mere speculation, but in 1998 a paper appeared that described a way in which the device could be made⁵. The author was Bruce Kane, then at the University of New South Wales (UNSW) in Sydney. Kane proposed burying phosphorus atoms 20 nanometres apart in a silicon grid. Each atom would have an electrode positioned above it. The qubits would be stored in the nuclear spins of the phosphorus atoms, and the voltage on the electrodes used to tune each nucleus to respond to a particular frequency of radio wave. In this way, the states of individual nuclei could be flipped using a radio signal sent to the chip. A second elec-

trode between adjacent phosphorus atoms would mediate the interaction between two neighbours, allowing operations to be performed using linked qubits.

The proposal has some distinct advantages. It would, for example, be possible in principle to build a structure containing thousands of qubits. And the qubits can be controlled by conventional electronics, which could, in theory, be easily integrated into the chip. "Even if there are some disadvantages to using silicon, the approach would probably win out over better technology because it would take advantage of existing infrastructure," says Stan Williams, director of quantum-science research at Hewlett-Packard Laboratories in Palo Alto, California.

Solid progress

With so much going for it, Kane's proposal was guaranteed attention. "There was a phase transition when the paper came out," says Henry Everitt, associate director of the physics division of the US Army Research Office in Durham, North Carolina, and chair of the US government's Quantum Information Science Coordination Group. "People who thought quantum computing was just a pipedream started thinking it could happen."

Robert Clark, a colleague of Kane's at UNSW, was one of the researchers to follow up the new idea. He is now director of Australia's Centre for Quantum Computer Technology, a multi-university initiative set up in 2000 that embraces about 150 researchers and that has made good progress towards making the electrodes and phosphorus grids that Kane's design requires.

There are two basic approaches to the construction of the device⁶. Bottom-up techniques use tools, such as a scanning tunnelling microscope (STM), that can manipulate individual atoms. STMs create atomic-resolution images by scanning a microscopic tip over a surface, but they can also manipulate individual atoms. The researchers use the STM to remove an atom from the layer of hydrogen that caps the surface of silicon wafers. When phosphine gas (PH₃) is blown over the wafer, the molecules drop into the gaps. The hydrogen can then be removed by heating the wafer. The team is currently trying to work out how to add a top layer of crystal-perfect silicon, which will be used to protect the phosphorus atoms.

Such manipulations represent the more



precise approach, but prototype devices are easier to make top-down. In this method, the Australian team covers the surface of a silicon wafer with a thin layer of a polymer containing an array of nanometre-sized holes. Next, the wafer is bombarded with phosphorus atoms. One atom shoots down each hole, punching its way about 15 nm into the silicon. The polymer around the hole is replaced and then carved away to make a stencil through which the circuits that control and read the qubit can be painted. "Many people looked at the proposal and said that the technology was decades away," recalls Gerard Milburn, a theoretical physicist at the University of Queensland and deputy director of the Centre for Quantum Computer Technology. "Well, I think we've proved them wrong."

Rather than working with nuclear spins, which are relatively difficult to manipulate, the Australian team is focusing on the outermost electrons of the phosphorus atoms. The ones and zeros are stored in the position of an electron that can hop between neighbouring phosphorus atoms. Clark's team is currently attempting to track the motion of the electron using a charge detector, and says that initial results are promising.

Under control

Researchers are impressed with the Australian team's progress. Adding phosphorus to silicon is common in the computer-chip industry, but no one knew how to do it one atom at a time when Kane made his proposal. "What they've done is spectacular," says Thomas Schenkel, a physicist at Lawrence Berkeley National Laboratory in California, who is working on a different implantation technique. "For the first time we can attempt to do electronics on a single-atom level."

Impressive, yes, but not yet quantum — the Australians have yet to obtain a working qubit. Other groups are also working on ideas stimulated by Kane's paper. Crispin Barnes, a semiconductor physicist at the University of Cambridge, UK, is working with colleagues to implant sodium atoms in silicon and store qubits in electron spin. And Kane, now at the University of Maryland in College Park, is working on techniques for measuring the spin of a single electron⁷.

The researchers pursuing these and other approaches believe that they will succeed, but there may be bigger problems ahead. Some believe that the electric fields associ-



Gerard Milburn: rapid progress is being made towards silicon-based quantum computers.

ated with the control apparatus could interfere with the qubits, disrupting the calculation. Others warn that slight variations between supposedly identical qubits, introduced by defects in the silicon or errors in aligning the electrodes, could render silicon quantum computers unworkable⁸.

Follow the light

This is where Stoneham's plan may have the edge. He proposes to manipulate the spins of electrons using laser light. In his design, the embedded atoms — he is currently considering which element to use — are arranged randomly in silicon, with no electrodes on top of them. Unlike the other silicon-based systems, which rely on ordered arrays, explains his colleague Andrew Fisher, Stoneham's design benefits from this disorder. The qubits are stored in the outermost electron on each atom. Because every atom has a unique set of neighbours, the qubit electrons each experience different forces and so can be manipulated individually using a specific frequency of laser light.

Linking the qubits together, however, is difficult. Stoneham plans to use 'control atoms' that are implanted across the wafer. As with the other embedded atoms, control atoms should be sensitive to a particular frequency of laser light. By tuning the pulse, the control atoms can be excited so that their electrons interact with the electrons that store qubits in neighbouring embedded atoms. This links nearby qubits, creating a two-qubit system. The states of the system, and hence the result of a calculation, are read by looking at the way in which photons scatter off the linked electrons.

Stoneham thinks that his processor should run at much higher temperatures than other silicon quantum computers. In

principle, other designs could work at up to 4 K. Above this, the electrons that carry the qubits will be disrupted by vibrations of the silicon lattice. But different atoms are sensitive to vibrations of different energies. Stoneham believes that some elements — perhaps bismuth — will bind their electrons in such a way that they are immune to the interfering vibrations. He is confident that his processor will work above 4 K, and maybe all the way up to room temperature. With the tools needed to build his processor already available, he is aiming to have a three-qubit system running in four years' time.

But that is a long way from a quantum processor with tens of bits, which Stoneham suggests he will have by 2010. He is a "brave man" to make such a claim, says Milburn. "I know I'm sticking my neck out," acknowledges Stoneham. But not everyone thinks he is being unrealistic. "The physics has not got far enough for us to say to Intel or Motorola: 'this is the chip we want you to build,'" says Everitt, "but that day may come in the next decade."

So what effect will Stoneham's ideas, and the other silicon devices, have on the race to build a practical quantum computer? The issue was addressed in a report published last year by the Advanced Research and Development Activity, a US government body that funds high-risk information technology research. It concluded that it is too soon to pick a winner, and that the ultimate technology may not have been invented yet.

Even Kane is hedging his bets. This June, the US Defense Advanced Research Projects Agency organized a meeting on quantum computing in Los Angeles, California. At it, Kane was asked whether the agency should initiate a quantum-computing version of the Manhattan project — the Second World War scheme that brought together high-level scientists and provided them with the investment needed to develop the first nuclear weapons. How did Kane respond? "I said no. We can't do that because we haven't even discovered fission yet," he says. ■

Jenny Hogan is a reporter for *New Scientist*.

1. Stoneham A. M., Fisher, A. J. & Greenland, P. T. *J. Phys. Condens. Matter* **15**, L447–L451 (2003).
2. Gulde, S. et al. *Nature* **421**, 48–50 (2003).
3. Jones, J. A. & Mosca, M. J. *Chem. Phys.* **109**, 1648–1653 (1998).
4. Vandersypen, L. M. K. et al. *Nature* **414**, 883–887 (2001).
5. Kane, B. E. *Nature* **393**, 133–137 (1998).
6. Clark, R. G. et al. *Phil. Trans. R. Soc. Lond. A* **361**, 1451–1471 (2003).
7. Kane, B. E. et al. *Phys. Rev. B* **61**, 2961–2972 (2000).
8. Keyes, R. W. *Appl. Phys. A* **76**, 737–741 (2003).

Impact factors reward and promote excellence

The system is unkind but effective. Others would do less good for developing countries.

Sir — I expected to read some robust criticism of Peter A. Lawrence's Commentary "The politics of publication" (*Nature* **422**, 259–261; 2003), so I was surprised at the chorus of approval in Correspondence (*Nature* **423**, 479–480 & 585; 2003, and *Nature* **424**, 14; 2003). These views and proposals require rebuttal. I believe that the present system of evaluation is the only one possible, and that Lawrence's apparently utopian proposals would do more harm than good.

It is a cliché that modern societies can hardly function without science, and that science has become very expensive and highly specialized, hence requiring an evaluation system. There are two socially justifiable reasons for supporting science. First, scientists make discoveries that increase our knowledge, understanding and predictive power. However, many well-educated people in all fields of science are needed to translate these into progress. Universities can produce these people and can test their skill and knowledge, but they cannot test the skill and knowledge of their own teachers, which has to be done through the engagement of the teachers themselves in scientific activity.

The second important reason for supporting science, therefore, is to teach students and to maintain a group of specialists in different fields who can adapt the newest scientific achievements to their society. Politicians and others who fund science need a tool to identify these people.

In the best laboratories, the first reason for maintaining science alone is considered paramount. But the second reason is vital to all modern societies, including those unable to produce Nobel prizewinners. Scientists maintain the polite fiction that all of them are equal and do equally good science. But this is not the case. The best laboratories make the most important scientific discoveries. A little lower are those in which less important discoveries are made, but which contain researchers who fully understand what others are doing and who can apply this knowledge. At the bottom are places where people only pretend to do science and are unable to follow progress in their field.

The system of rewards in science must assure promotion of the best laboratories, improvement of the decent and denial of public funds to the worst. Neither international congresses nor big international programmes can make this objective distinction between good and poor science, so some other means of evaluation are required.

The appearance of the Science Citation Index (SCI) in the 1960s was a breakthrough in the development of objective numerical methods for the evaluation of science and scientists. This can be seen by comparing now with then, and by looking at places where numerical methods of evaluation are unknown. In countries far behind the scientific leaders, scientists are no less numerous, and many universities and scientific journals are supported by public funds. These journals publish many papers but have very low circulations and an insignificant impact on other scientists. This is a waste for the society supporting such research, as the scientists cannot make important discoveries, convey or build on discoveries made by others, or follow developments in their own field.

Sometimes this can be seen in rich countries too. In the 1960s and 1970s it was a waste of time to browse German and French journals on ecology and evolutionary biology. This state of affairs changed completely after young researchers started to be rewarded for publishing in journals with a high impact factor: now German and French researchers in these fields write papers that are well worth reading.

Evaluation of scientists on the basis of the impact factor and other indices is like the market economy: the system is wrong and unjust, but other systems are much worse. Thousands of books have been written on the evils of capitalism, and now we have articles on the evils of evaluations derived from citation indices. The authors

of these articles ignore the global effect of applying this system and concentrate instead on particular cases: a paper got many citations despite being published in a journal with a low impact factor, or a poor paper was cited many times. Evaluation based on citations is a statistical method that has to be used on large samples and carefully applied to avoid pitfalls. Arguments against the system should be statistical, not particular.

Critics of this evaluation system propose a utopia with high moral standards. They want science managers and journal editors not to be narrow specialists but to be able to evaluate scientists in all different fields within, say, ecology or molecular biology. With the present extent of specialization this seems hardly possible. In this utopia, managers and editors would be absolutely honest and not guided by their own scientific interests, predilections or aversions. The entire system relies on the best side of human nature.

Abandonment of objective methods of science evaluation derived from the SCI would be most dangerous in developing countries and others where science is not first-rate. It would keep their societies from knowing how far behind their scientific institutions are. Worse, it would remove a tool for rewarding researchers who attempt to do good science and for eliminating those who do not.

Adam Łomnicki

*Institute of Environmental Sciences,
Jagiellonian University, ul. Ingardena 6,
30-060 Kraków, Poland*

Impact factors: letting everyone have their say

Sir — Am I the only person to have been struck by the apparent contradiction between the space you give on the one hand to those critical of the impact of impact factors — for example Peter Lawrence's excellent Commentary piece "The politics of publication" (*Nature* **422**, 259–261; 2003) and the subsequent correspondence (*Nature* **423**, 479–480 & 585; 2003) — and on the other hand your advertising department's decision to fill two whole pages with a striking example of what Adrian Tuck in Correspondence (*Nature* **424**, 14; 2003) rightly calls "bragging": the news of *Nature's* 30.432 (note the three decimal places!) impact factor for 2002?

Perhaps *Nature* should speak with a

single editorial voice; otherwise some of your readers might begin to regard your coverage of impact factors with a slight degree of cynicism.

Matthew Cobb

*School of Biological Sciences,
University of Manchester, Oxford Road,
Manchester M13 9PT, UK*

Impact factors are but one measure of a journal's performance. We are proud of our achievements on this measure. The debate, which we have long encouraged, is about how impact factors are used — Editor, *Nature*.

correspondence

Contributions to Correspondence may be submitted to corres@nature.com. They should be no longer than 500 words, and ideally shorter. Published contributions are edited.

Beware of false profits

Academics must consider the true cost of dealings with the marketplace.

**Universities in the Marketplace:
The Commercialization of Higher
Education**

by Derek Bok

Princeton University Press: 2003. 256 pp.

\$22.95, £15.95

Jeremy Gunawardena

The president of a leading university uses a US\$2-billion loan to attract "world-renowned scholars, dazzling new buildings, incredibly talented students". The president is fêted as an innovative academic leader, but the inexorable demands for repayment of the loan begin a steady process of decline. The athletic programme is refurbished to raise lucrative television income. The university museum is forced to divest itself of its treasures. The rights to exploit the university's biological research are sold to a pharmaceutical company. A for-profit Internet venture is launched selling degrees to students worldwide. Advertisements are placed in classrooms, and company logos appear on course material. Finally, in desperation, 100 undergraduate places are auctioned to the highest bidder.

This not-so-hypothetical nightmare must have surprised the audience at Harvard University's 1988 commencement when Harvard's then president, Derek Bok, recounted it to drive home his concern about the perils of commercialization. The nightmare has continued to haunt him, and he returns to it in this humane and beautifully crafted book. Bok believes that the intrusion of the marketplace into the university is eroding fundamental academic values, and that we must act now to halt this decline.

The universities and marketplace discussed by Bok are American, and he has nothing to say about the very different histories and contexts elsewhere. But readers who are puzzled by the attention given to university sports, a peculiarly American obsession, will find much here that is of universal interest.

Corporate sponsorship of research is a concern for many readers of *Nature*. Bok recounts the cases of two researchers whose attempts to publish the results of negative clinical trials elicited ugly harassment from the drug companies involved but little support from their respective institutions. Such egregious examples at least attract considerable attention. The subtle, and sometimes not so subtle, effects of distortion and undue influence are much more damaging when they are largely invisible. A survey of academic studies on the effects of passive smoking found that, of the investigators with ties to the tobacco industry, 94% claimed that it is not



**Crust under foot? Protesters failed to stop
Novartis president Douglas Watson giving \$25
million to the University of California, Berkeley.**

harmful, whereas 87% of those without such ties claimed that it is. Conflicts of interest that would be considered scandalous in public life elicit little comment, until some unfortunate incident attracts public attention.

As Bok argues, the remedy for such ills is to enforce full disclosure of all forms of sponsorship. It is troubling that such an obvious suggestion, so consistent with academic values, has not already been implemented by more universities and journals (*Nature* adopted such a policy in 2001; see *Nature* 412, 751). Are some academics concerned that the uninformed public may draw the 'wrong' conclusions from such revelations? If so, the public will eventually draw the right conclusions about such academics.

The commercialization of teaching raises a different set of issues. Bok recognizes the potential of the Internet to enrich the learning experience, in the face of the faculty's resistance to change (a deeply cherished 'academic value'). As Bok once said: "I fully intend to break my lance, as presidents have before me, against the wall of better teaching." Here he attempts to balance the wall-breaking potential against the dangers of selling poor-quality courses to ill-informed students for a quick, and probably illusory, profit.

It is a difficult balancing act throughout. The graceful language and judicious argu-

ments do not conceal the leaps that Bok has to take in navigating across this ethical quagmire. Private universities charge fees. Business and medical schools run their professional programmes as lucrative, profit-making enterprises. The money may not go into shareholder's pockets, but that may be of little comfort to those who pay. If commercial companies insist on secrecy, it is nothing compared to the lengths to which some academics go to shield their ideas from their own colleagues. So graceful is Bok under such pressures that the odd hesitation goes almost unnoticed. The Novartis Department of Plant and Microbial Biology — I beg your pardon, the University of California, Berkeley, Department of Plant and Microbial Biology and its \$25-million deal with Novartis — was "probably not a significant threat to academic values".

Bok is acutely sensitive to the complexities of institutional governance and to the incentives that drive individuals. He is less attuned to the cultural dynamic. Science and technology are no longer independent, the former to be undertaken for its own sake and the latter for profit. They have become symbiotic. Indeed, in some cases — quantum computing comes to mind — industry, not academia, has made the running. It is not profit-making that drives universities and companies to jostle on the intellectual commons; it is the jostling that drives the profit-making. As the example of quantum computing suggests, the pattern of interaction between academia and industry varies from discipline to discipline. If the biosciences need lines drawn to prevent abuses, the physical and computational sciences need encouragement to foster more collaboration between academia and industry, not less.

As Bok points out, there is a long history of gloomy prophecies about the decline of the university. None of these have come to pass. Here too there is a powerful dynamic at work. The intellectual life is a vocation, and the university has become its only real haven. If some academics and universities make frightful blunders, others will learn from them — as many will do from this thoughtful and thought-provoking book. ■

Jeremy Gunawardena is at the Bauer Centre for Genomics Research, Harvard University, Cambridge, Massachusetts 02139, USA.

More on US universities

Who's in Charge of America's Research Universities? A Blueprint for Reform

by Thomas J. Tighe

State University of New York Press, \$59.50 (hbk), 19.95 (pbk)

A child in time

Portrait of the Artist as a Child: The Gravettian Human Skeleton from the Abrigo do Lagar Velho and its Archeological Context edited by João Zilhão & Erik Trinkaus
Instituto Português de Arqueologia: 2003. 610 pp. €40; distributed by Oxbow Books, £35, \$65

Paul Bahn

The discovery in Portugal of the 'Lapedo child' in December 1998 was an important event for those studying European pre-history. For a start, this was the first Ice Age burial ever discovered in the Iberian Peninsula. Second, the skeleton was largely intact, although it had been disturbed by a bulldozer about ten years ago. Third, it was excavated with the most meticulous and exemplary care by Cidália Duarte and others. And finally, the child quickly became famous, even notorious, around the world because of claims that it might be a hybrid of Neanderthals and Cro-Magnons.

Portrait of the Artist as a Child is a fine monograph, beautifully produced in English, that presents scholars with a huge quantity of basic data. The first 200 pages cover the archaeological context of the burial. Although the human remains themselves could not be dated, the animal bones and charcoal found with them place the burial firmly at between 24,000 and 25,000 years ago, in the Gravettian period.

There is evidence that a shallow pit was

dug in the back wall of a rock shelter and a branch of Scots pine was burned at the bottom. The child was then placed in the pit. Red-ochre stains on both the upper and lower surfaces of the bones, and a clear boundary with the surrounding whitish sediment, suggest strongly that the body was wrapped in an ochre-painted shroud. It lay in an extended position, with a young dead rabbit placed across its lower legs, and with the pelvises of two red deer (perhaps meat offerings) by its shoulder and its feet. Round its neck was a perforated-shell pendant, and on its forehead was some kind of head-dress made up of four canine teeth from two different red-deer stags and two hinds. This burial can therefore be grouped with roughly similar Gravettian ochre burials known in Britain (at Paviland), Russia (Sungir) and especially Moravia.

Most of the book — no less than 275 pages — is devoted to an exhaustive inventory of the human bones and teeth, making available to the scientific community all of the morphological details and measurements. Different methods of determining dental age at death give a mean age of 4.7 years, and the skeletal age at death corresponds with this assessment.

From the start it was clear that this was an anatomically modern child — it has a chin and other modern features — but the bodily proportions, especially those of the legs, suggest some Neanderthal input. This claim led to an academic furore that smacked more of an argument based on emotion and entrenched beliefs than of an open-minded, cold and logical debate.

This monograph should go a long way towards resolving the issue, although those with extreme opinions are unlikely to be swayed. The data seem to have been provided as objectively as possible; for example, the morphology of the child's bony labyrinth (inner ear) points to it being a modern human, yet the authors of that section say that their data are inconclusive with regard to the hybrid theory.

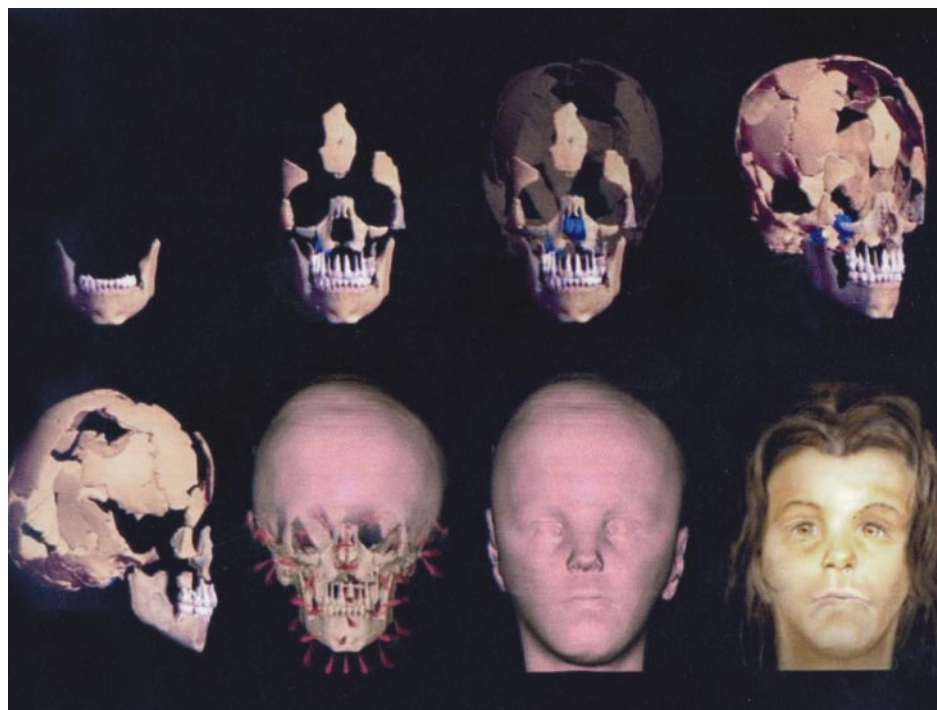
Overall, the Lapedo child is clearly not a normal anatomically modern human, and it displays an unusual mosaic of postcranial characteristics — especially in the lower limbs, but also in some features of the upper ones. It also has a pitting of the occipital bone in the skull, which some specialists consider virtually diagnostic of Neanderthals.

The editors (who are also the principal investigators of the child) do consider possible alternative explanations — such as nutritional or climatic factors — for the child's unusual proportions. But they conclude that no alternative hypothesis has yet been presented that fits the data better than their hybridization theory.

The book ends with some remarkable and useful surveys of Neanderthal and early Upper Palaeolithic burials of different kinds, and of the whole contentious issue of the Middle-to-Upper Palaeolithic transition. Most readers, however, will probably only want to know if this is just a "chunky modern kid", as it has sometimes been described, or whether it is really a hybrid of Neanderthals and Cro-Magnons. The question remains open, and the editors have taken pains to present all the data, which others can use to form an opinion, while arguing very persuasively in favour of their own view.

What's crucial is that the issue should be presented accurately. This is emphatically not a question of whether any evidence of Neanderthal ancestry can be found in present-day Europeans, but of whether such evidence can be found in early Upper Palaeolithic Europeans. The much-touted DNA analyses carried out in recent years on a few Neanderthal bones are therefore of no consequence here whatsoever. Nobody is claiming Neanderthal genetic continuity to the present day — descendants are irrelevant, and in any case this child almost certainly had none. It is a question of ancestry, and, faced with the evidence assembled here, it is hard to deny the strong possibility that this child could indeed have had mixed antecedents. ■

Paul G. Bahn is the consultant editor of *Written in Bones* (Firefly, \$35 (hbk), \$24.95 (pbk)).



Adding flesh to the bones: studies of the Lapedo child's skull reveal both Neanderthal and Cro-Magnon features.

Nature's magic bullets

Antibiotics: Actions, Origins, Resistance

by Christopher Walsh

ASM Press: 2003. 335 pp. \$99.95; \$89.95 to members of the American Society for Microbiology

Stewart T. Cole

Starter units, expansion, modification, mergers, restructuring, down-sizing and cleavage. There's nothing new there — today's pharmaceutical industry is merely following the business plan laid down by nature for the production of antibiotics by microbes. In the face of market forces, the creation of giant pharmaceutical companies that concentrate extensive resources on lifestyle diseases, at the expense of the research and development (R&D) of new anti-infective agents, is giving justifiable grounds for concern. In this book, devoted exclusively to antibacterial agents, Chris Walsh provides an authoritative account of their (bio)synthesis, mode of action and resistance mechanisms, and highlights new strategies to find and improve antibiotics. This latter topic should prove to be a source of inspiration for the R&D directors of big pharmaceutical companies.

The first of the book's five sections outlines the concepts to be developed later. It highlights the slow pace of development of new agents in recent decades. The second section introduces the main classes of antibiotic that target cell-wall biosynthesis, the metabolism of nucleic acids, the protein-producing machinery and the folic acid pathway. Like the rest of the book, it is superbly illustrated: several colour plates take full advantage of the remarkable progress in structural biology to show, for instance, how macrolide antibiotics bind to the polypeptide exit tunnel of the ribosome and prevent proteins from being made.

The third section concentrates on the principal resistance mechanisms that block the action of antibiotics on pathogens and relates how these are commonly used by the antibiotic-producing organisms to protect themselves from their own products. These mechanisms include modifying the target, destroying the antibiotic, and lowering its concentration to ineffective levels. The importance of active efflux to antibiotic resistance is also given due attention, although the trick of altering cell permeability to reduce the entry of antibiotics is only addressed in a cursory manner. Parts of the book would have benefited from critical appraisal by a microbiologist, but this is a minor quibble.

In a fourth section that is somewhat

Online database

Island investigations

More than a hundred European scientific expeditions went to the Canary Islands during the eighteenth and nineteenth centuries. Many were on their way to more distant destinations — but others were there to map and study the natural history of the islands themselves, and led to many scientific discoveries of general importance.

The original reports of these expeditions, physically scattered around the world, are now being brought together courtesy of a major European Commission-funded digital project known as ECHO (European Cultural Heritage Online).

Some of the works have already been widely seen. The 1856 expedition of Charles Piazz-Smith, for example, was recorded in his popular work *An Astronomer's Experiment* (1858), held in several libraries in Spain and Britain. Piazz-Smith made astronomical observations at different altitudes on El Teide, Tenerife's highest peak, showing for the first time that more stars can be seen at the top of a mountain.

By contrast, the work of the Norwegian



botanist Christen Smith, who accompanied German volcanologist Leopold von Buch in his 1814 expedition, was thought to be lost, as Smith continued on to the Congo where he died. But his manuscripts and diaries, describing some 600 different plants, about 50 of which were new to science at the time, were recently found in a library in Oslo.

The picture shows illustrations of volcanoes taken from *Atlas des Isles Canaries* (1836) by von Buch, who developed his theory of volcanoes in the Canaries.

Alison Abbott

<http://humboldt.mpiwg-berlin.mpg.de>

anthropocentric in its reasoning, we are introduced to the molecular rationale that governs the assembly-line production by actinomycetes and fungi of antibiotics belonging to the polyketide and non-ribosomal peptide classes. We encounter cascades of enzymes and massive protein factories that, following the active growth phase, manufacture secondary metabolites, many of which act as antimicrobial agents or immunosuppressors. The environmental cues and regulatory effects of the signal-transduction networks that govern the expression of the corresponding genes are also explained, but insufficient attention is paid to the forces of natural selection. However, Walsh lucidly expounds how genomics, biochemistry and structure determination allow one to predict the composition of a polyketide if you know the arrangement of the catalytic domains of the enzyme responsible for its synthesis; he implies that the reverse prediction is also feasible.

This sets the tone for the fifth and final section, which elegantly relates how new antibiotics can be produced from existing compounds or biosynthetic pathways. It also explains why engineering polyketide synthases, and other enzymes that produce or modify antibiotics, will enable us to generate in a rational manner new agents that could have therapeutic value.

The strength of the book lies in Walsh's mastery of chemistry and his ability to explain complex (bio)chemical reactions in a simple yet vivid fashion. The reader will

conclude, correctly, that most of our clinically useful antibiotics stem from research on microbes, and that, with just a few exceptions, synthetic and combinatorial medicinal chemistry have yet to achieve their full potential. Nonetheless, Paul Ehrlich, Alexander Fleming and Selman Waksman would all have found something to enjoy in *Antibiotics*. ■

Stewart Cole is in the *Unité de Génétique Moléculaire Bactérienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris, France.*

Light summer reading

Isaac Newton

by James Gleick

Pantheon, £22 (also Fourth Estate, £15, available in September)

An acclaimed new biography of one of the fathers of physics.

Einstein

by P. D. Smith

Haus, £12.99 (hbk), £8.99 (pbk)

A new concise biography of the famous physicist.

Lovelock and Gaia

by Jon Turney

Icon, £9.99

A brief account of James Lovelock's vision of a self-regulating planet and his influence on Earth system science.

Skeleton keys

What was your first experiment as a child?

I have been fascinated by skeletons since the age of six. One of my early adventures was to unearth a medieval human skull from a ditch near an old church. As I was brought home by a policeman with the skull under my arm, obstinately refusing to put it back, my father agreed with the policeman that I could keep the skull for a while and then return it to its eternal rest. I cleaned the skull, restored it, and kept it for many years. Then I began to mount skeletons of animals I found dead on the roadside. This made the family kitchen look like an autopsy lab.

One day, I discovered an abandoned slaughtering-place in the forest. Here I unearthed century-old carcasses of horses, cows, goats, donkeys and so on (I was blessed with much room at home, and ignorance of anthrax). By mounting so many skeletons, I discovered that what attracted me was, essentially, homology. Nothing was more satisfying to me than discovering that, for example, the wishbone was the clavicle. Later, I turned to the fossil vertebrates of my native Touraine. Fragmentary bones of fish, crocodiles, rhinos and other fossil mammals revealed more amazing homologies. Much later, the same motivation led me to tackle the anatomy of weird and far older fish.

What book has been most influential in your scientific career?

Systematics and Biogeography, Cladistics and Vicariance by Gareth Nelson and Norman Platnick was like a cold shower and made me think in an entirely different way.

What gives you the most job satisfaction now? What are your frustrations?

The supreme delight is to find, by chance, something entirely unexpected. In this respect, field work in the Palaeozoic rocks of South America is a favourite source of satisfaction. My frustration lies in not being fully able to understand all the exciting papers in developmental genetics that come out these days.

What was the most memorable comment you ever received from a referee?

By some mischance a copy of one of my manuscripts contained the same page twice, yet one of these pages was placed far from its twin. For one paragraph of the first of these exactly similar pages, a referee wrote in the margin "yes, excellent", and for the same paragraph of the other page: "remove this".

Where and when would you most like to have lived or worked?

France, in the second half of the eighteenth century.

Before or after 1789?

Both. I should have liked to have been a middle-class provincial (perhaps a physician) in the time of the enlightenment, then experience the revolution and, notwithstanding the risk of experiencing the guillotine, die peacefully before the fall of 'the dwarf'.

What book currently resides on your bedside table?

The translation of Gregory of Tours' *Historia Francorum*, which I often re-read when I am worried by problems in scientific politics. It makes me realize that things could be worse.

What literary character would you employ as a postdoc?

Jules Verne's Captain Nemo.

Assuming the dead can be raised and/or time travel exists, who from the world outside science would you most like to have dinner with?

Aristotle (notwithstanding the probable frugality of the dinner).

You are on a plane behind two students obviously going to the same conference, who start to talk about your work. What do you do?

Insert ear-plugs and say nothing.

What one thing would you rescue from your burning laboratory?

My copy of the *Handbuch der Vergleichenden Anatomie der Wirbeltiere*, edited by Bolz, Göppert, Kallius and Lubosch (1931–1937). It is an incomparable source of data, not otherwise readily available, and would be enough for mulling over vertebrate phylogeny for years on a barren island.

What would you have become, if not a scientist?
A chef (note: I have changed since I boiled decaying hedgehogs in my parents' kitchen).

What's your motto?

Never answer "yes" immediately.

What previously under-recognized sport or pastime should be included in the Olympic Games?

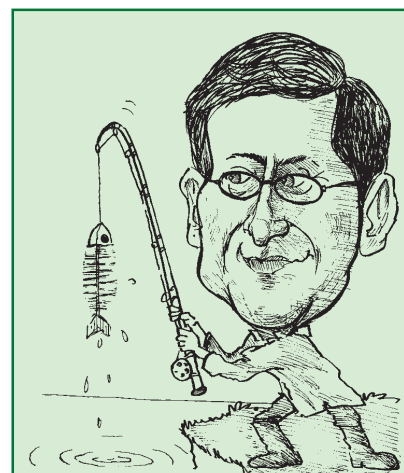
Hagfish-catching (with hands only).

What single discovery, invention or innovation would most improve your life?

A machine for expanding space (especially houses and flats) without limit.

What overlooked or underrated discovery really changed the science in which you work?

The use of acetic acid to etch fossils from limestone.



Philippe Janvier

Philippe Janvier is a palaeontologist at the CNRS and works at the Muséum National d'Histoire Naturelle in Paris, France. He is also an honorary research fellow at the Natural History Museum in London. He likes angling, cooking and loves eighteenth-century books.

You've just been told (in confidence) that the world will end tomorrow. What would you do next?

I would have fried gudgeon with Muscadet, then three-dozen snails (the small species *Helix aspersa*, not the big, Burgundy one) with a Cravant 2000, and finally goat's cheese (preferably an old, hard Sainte-Maure) with a Bourgueil 1996. (The gudgeon and snails caught and prepared by myself — part of the pleasure.)

What's the most interesting thing in your fridge?

A can of Swedish *surströmming* (fermented Baltic herring), which will soon be out of date.

Do you have a burning ambition to do something of no practical or immediate value?

Restore a huge (but really huge) castle. Apart from that, I'd like to go to a rave party (just once, to see what it was like). I like techno music in small doses — I think it is a real innovation that could lead to an entirely new music. The great step will occur with the composition of 12-tone techno (perhaps using Schönberg or Berg as a base).

What music would you have played at your funeral?

Messiaen's *The Lord's Nativity*. It is admittedly a bit long and might be more suitable for a cremation. But, as a palaeontologist, I simply cannot accept cremation (this would ruin the profession).

The host at a dinner party has placed you next to God. What would you talk about?

Finalism.

A dose of the bomb

Mark P. Little

Measurements of high-energy neutron exposure in Hiroshima validate estimates of the amount of radiation that survivors of the atom bomb received. Can we now predict the risks of radiation more reliably?

On page 539 of this issue, Straume *et al.*¹ present new measurements of the levels of neutron radiation to which survivors of the Hiroshima atomic bomb were exposed 58 years ago. Their calculations resolve a controversy that has existed for two decades — were previous estimates of neutron exposure accurate? It turns out that the answer is yes. But why is this important? We are all exposed to ionizing radiation (X-rays, γ -rays, α -particles and neutrons) — in hospitals and dental surgeries, at home or, in some cases, at work (see Box 1; ref. 2). Health data collected from survivors of the atomic bombs in Hiroshima and Nagasaki are the basis for most estimates of the risks associated with ionizing radiation, which in turn are used to set safe limits for radiation exposure^{2–4}. But our ability to estimate risk from this data set is underpinned by our ability to correlate the incidence of disease with the dose of radiation received. If we do not know the radiation dose, we cannot predict the risks.

The atomic bomb, dropped on 6 August 1945, destroyed Hiroshima (Fig. 1). Survivors of the explosion were exposed to two types of radiation: γ -rays and neutrons. There have been several estimates of how much radiation they received, based on interviews with survivors and responses to questionnaires about their location at the time of the bombings. The most recent set of estimates — the so-called DS86 dosimetry



Figure 1 **Hiroshima, October 1945.** This scene of devastation is the Kamiyacho Road, near the Electric Building, facing west towards the centre of the explosion. (Photo supplied by the Library and Archives Section, Archives Office, Department of Information Technology, RERF.)

system — was introduced in 1986 (ref. 5). But even this dosimetry system was known to be associated with serious problems over the neutron-dose estimates for Hiroshima. Neutron radiation emitted by the atomic bomb consisted of both high-energy 'fast' neutrons and low-energy 'thermal' neutrons — and DS86 predicted that there should

have been less low-energy neutron exposure at distances of more than 1,000 metres from the bomb than actual measurements at that distance had indicated^{5–7}. Although low-energy neutron radiation represented only a tiny fraction of the total neutron radiation that survivors received, the discrepancy between DS86 and the low-energy neutron measurements had troubling implications for the dosimetry system as a whole. It cast doubt on the accuracy of estimates of radiation levels emitted by the Hiroshima bomb, and also on the calculations of radiation transport through the air^{5,8}.

Previous measurements of high-energy neutrons had been unable to verify the accuracy of DS86 at these distances. Until a few years ago, such neutrons could be measured only in sulphur samples. Sulphur measurements made close to the centre of the explosion agreed reasonably well with the DS86 estimations, but there were no useful data more than 1,000 metres from the bomb⁵.

Straume *et al.*¹ have now measured high-energy neutron exposure over the distances (900–1,500 metres) where the DS86 estimations differed from the low-energy neutron measurements — the range also relevant to survivor locations. The authors used

Box 1 Sources of ionizing radiation

All humans are continually exposed to ionizing radiation, mostly from natural sources. Worldwide, the average human exposure to radiation from natural sources is 2.4 millisieverts (mSv) per year. About half of our exposure comes from the radioactive decay products (mostly α -particles) of radon gas. Most of the remaining natural exposure comes from cosmic rays and terrestrial γ -rays. At ground level, most of the dose from cosmic rays comes from muons, but at aircraft

altitudes, neutrons, electrons, positrons, photons and protons provide most of the dose. Dose increases with altitude, so 100 hours of air travel (such as a frequent flyer might travel in one year) would add an extra 0.5 mSv.

Of man-made sources of radiation, most exposure comes from diagnostic medical procedures such as X-rays, and these contribute about 0.4 mSv per year on average. More than 500 atmospheric nuclear explosions, including those at Hiroshima and

Nagasaki, took place between 1945 and 1980. Continuing fallout from these gives an annual average dose of around 0.005 mSv. The Chernobyl accident, which occurred in 1986, adds about 0.002 mSv per year on average, and nuclear power production about 0.0002 mSv per year. To put these figures in perspective, the average dose to the colon for the Japanese atomic-bomb survivors was 200 mSv, with the maximum dose in excess of 5,000 mSv (ref. 19).

M.P.L.

technology that has been developed to detect trace amounts of the long-lived nickel radioisotope ^{63}Ni , which is produced from copper atoms by high-energy neutrons. Straume and colleagues have detected informative quantities of ^{63}Ni in copper samples taken as far away as 1,500 metres from the Hiroshima bomb, and their new calculations of neutron exposure agree with the DS86 estimations over these wide distances.

Closer to the centre of the explosion (around 380 metres), the DS86 estimations may indeed have been wrong. The ^{63}Ni measurements suggest that actual exposure at this distance was around 35% less than predicted, a finding that is supported by the earlier measurements in the sulphur samples⁵. But because most of the Hiroshima survivors who suffered appreciable exposure were between 900 and 1,700 metres from the explosion⁵, this discrepancy is of largely academic concern.

Taken together with previous validations of the estimated γ -ray doses⁵, and of the Nagasaki neutron doses⁹, it is now clear that the DS86 dose estimates correctly reflect all components of radiation dose in the two cities. So where does this leave the measurements of the low-energy neutrons that initiated these investigations? Embarrassingly, recent re-analyses of these data suggest that if background radiation is taken into account (which the original analyses did not do), then the discrepancy with DS86 largely disappears^{10,11}.

What are the implications of the new study for risk estimates? One implication is clear: Straume *et al.*¹ have confirmed that neutrons accounted for only 1–2% of the total radiation dose received by the survivors of the atomic bomb (although after accounting for their greater biological effectiveness relative to γ -rays², the proportion of the total dose becomes 10–20%), so we cannot derive any useful information about the risks associated with neutron exposure from the Hiroshima bombing^{12,13}. But on the positive side, we now have more confidence both in the estimates of the total radiation output from the Hiroshima bomb, and in the calculations of radiation transport through the air. We also have greater confidence in determinations of the relative proportions of γ -ray and neutron radiation to which the survivors were exposed. Will predictions of the risks associated with γ -radiation therefore become more reliable? Arguably yes, but the atomic-bomb data remain contentious as a source of risk estimates for reasons that are largely independent of the dosimetry. Stewart and Kneale¹⁴ maintain that the survivors of the Hiroshima and Nagasaki bombings are highly 'selected', with the degree of selection depending on age as well as the dose of radiation. They argue that such selection invalidates the use of the survivor data for

deriving risk estimates that are applicable to a general population.

Is this a real problem? The work of Straume *et al.*¹ and others^{5,9} indicates that there are unlikely to be appreciable systematic inaccuracies in the DS86 dose estimates; however, random errors in the dose estimates for individuals still exist. For example, the bomb survivors' recall of their position in the two cities at the times of the bombings was inevitably imprecise. Interestingly, if these random individual errors are taken into account, the selection findings of Stewart and Kneale¹⁴ largely disappear¹⁵. Moreover, the cancer risks derived from survivor data are statistically consistent with those observed in groups exposed occupationally^{16,17} and medically^{2,18}. So, despite individual errors, the collective data from the survivors of the atomic bomb are likely to remain a valuable predictor of the risks of ionizing radiation. ■

Mark P. Little is in the Department of Epidemiology and Public Health, Imperial College Faculty of Medicine, St Mary's Campus,

Norfolk Place, London W2 1PG, UK.

e-mail: mark.little@imperial.ac.uk

1. Straume, T. *et al.* *Nature* **424**, 539–542 (2003).
2. Sources and Effects of Ionizing Radiation. *UNSCEAR 2000 Rep.* Vols I, II (UNSCEAR, New York, 2000).
3. ICRP Ann. *ICRP* **21** (1–3) (1991).
4. NRPB Docs *NRPB* **4** (4), 15–157 (1993).
5. Roesch, W. C. (ed.) *US–Japan Joint Reassessment of Atomic Bomb Radiation Dosimetry in Hiroshima and Nagasaki* (Radiation Effects Res. Found., Hiroshima, 1987).
6. Straume, T. *et al.* *Health Phys.* **63**, 421–426 (1992).
7. Shizuma, K. *et al.* *Health Phys.* **65**, 272–282 (1993).
8. Gold, R. *Radiat. Meas.* **24**, 9–29 (1995).
9. Straume, T., Harris, L. J., Marchetti, A. A. & Egbert, S. D. *Radiat. Res.* **138**, 193–200 (1994).
10. Status of the Dosimetry for the Radiation Effects Research Foundation (DS86) (Natl Acad. Press, Washington DC, 2001).
11. Huber, T., Rühm, W., Hoshi, M., Egbert, S. D. & Nolte, E. *Radiat. Environ. Biophys.* **42**, 27–32 (2003).
12. Little, M. P. *Int. J. Radiat. Biol.* **72**, 715–726 (1997).
13. Hunter, N. & Charles, M. W. *J. Radiol. Protection* **22**, 357–370 (2002).
14. Stewart, A. M. & Kneale, G. W. *Int. J. Epidemiol.* **29**, 708–714 (2000).
15. Little, M. P. *Int. J. Radiat. Biol.* **78**, 1001–1010 (2002).
16. Cardis, E. *et al.* *Radiat. Res.* **142**, 117–132 (1995).
17. Muirhead, C. R. *et al.* *J. Radiol. Protection* **19**, 3–26 (1999).
18. Little, M. P. *Int. J. Radiat. Biol.* **77**, 431–464; 745–760 (2001).
19. Pierce, D. A., Shimizu, Y., Preston, D. L., Vaeth, M. & Mabuchi, K. *Radiat. Res.* **146**, 1–27 (1996).

Global change

South dials north

Thomas F. Stocker

Climate is greatly influenced by ocean circulation in the North Atlantic. But warming episodes, as glacial conditions turned into interglacials, may have been triggered by events far to the south.

The Earth's emergence from the grip of the last ice age took about 10,000 years and was a rough ride. Cycles of rapid warming and cooling preceded a final period of warming, and entry into the current interglacial, from about 10,000 years ago. The dominant player in these events is thought to have been the North Atlantic thermohaline circulation, which transports massive amounts of heat northwards from the tropics, and the effect on that circulation of perturbations in the North Atlantic itself¹. But might the crucial dials that control this system be elsewhere? On page 532 of this issue, Knorr and Lohmann² present a modelling study which shows that slow climate changes in the Southern Ocean around Antarctica (Fig. 1) can influence events in the North Atlantic. Those changes, it seems, may have been ultimately responsible for the abrupt warmings recorded in the Northern Hemisphere.

The Atlantic thermohaline circulation is mainly driven by density differences in bodies of water, density being determined by temperature and salinity. Warm water flows north from the tropics, cooling and sinking as it approaches high latitudes to become a return flow at depth as North Atlantic Deep Water³. The traditional view, based on

studies of how this 'meridional circulation' can collapse, is that bursts of fresh, less dense water from melting northern ice sheets or background noise in the North Atlantic trigger instabilities that cause it to weaken or shut down^{4,5}.

Knorr and Lohmann² have looked instead at how this circulation can resume after being stalled. They find that, once a threshold is reached, slowly increasing sea surface temperatures around Antarctica and receding sea-ice cover lead to the North Atlantic thermohaline circulation being rapidly switched on. The abrupt increase of meridional heat transport by the ocean causes a sea surface warming of up to 6 °C in the northern North Atlantic within a few decades. These results constitute significant progress in climate studies. They show that slow changes in the south can have abrupt and far-distant consequences, and they link processes operating on scales of thousands of years (such as alterations in Earth's orbital parameters) with the faster changes occurring in, for example, the thermohaline circulation or ice-sheet discharges.

Knorr and Lohmann's climate model is a comparatively simplified one which is efficient for investigating processes associated with deep-ocean circulation and its long

adjustment time. But it largely omits ocean interactions with atmospheric dynamics, particularly with the hydrological cycle. As in many models, the Atlantic thermohaline circulation exhibits hysteresis behaviour when affected by freshwater fluxes — that is, the route taken to an end point is not the same as the return route to the starting point, and this property is at the heart of the mechanism described by Knorr and Lohmann. Here, hysteresis implies that, within a certain range of heat and freshwater exchange between the atmosphere and ocean, the thermohaline circulation can be either weak or strong.

In the authors' experiments, ocean surface parameters south of 30° S are modified slowly from those of glacial conditions to those characteristic of the present interglacial (sea surface temperatures increase and sea ice retreats). Both effects decrease the density of surface waters and, because of Earth's rotation, the Antarctic Circumpolar Current accelerates. This increases the return flow of near-surface waters into the South Atlantic via paths that affect the salt balance of the Atlantic basin⁶. Northward motion of water masses also increases in the North Atlantic, bringing denser, salty waters from the tropics into the 'sinking' regions at high latitudes, where they eventually kick-start thermohaline circulation and meridional transport of heat.

Another mechanism of distant triggering of thermohaline circulation in the Atlantic has been proposed^{7,8}, invoking injection of fresh water from the Antarctic ice sheet. In consequence, the density of deep and intermediate waters from Antarctica, which compete with North Atlantic Deep Water, is reduced, ultimately accelerating the thermohaline circulation⁹. This mechanism



Figure 1 **Southern perspective: a satellite view of Antarctica and the Southern Ocean.** Knorr and Lohmann's study² suggests that events in this region could have triggered warming in the north.

may also be active in Knorr and Lohmann's simulations: in their model, at 30° S, most of the Atlantic Ocean below a depth of 500 m becomes fresher, and hence less dense, when the surface conditions around Antarctica are forced to change.

The North Atlantic thermohaline circulation is known to be vulnerable to pulses of melt water into, or close to, the areas where deep water forms at high latitudes, and this can shut down deep-water formation⁴, with possible global implications through the operation of the 'bipolar seesaw'¹⁰. In this mechanism, an abrupt cooling in the north, caused by the shut-down of the thermohaline circulation, would initiate a slow warming in the south. Given these over-

all results^{2,7,8}, however, it seems that reactivation of that circulation is likely to stem from events in the far south: by southern warming and increased northward advection of salt at the surface; or by a reduction in the density of deep waters of southern origin entering the Atlantic Ocean at depth, so giving way to enhanced formation of North Atlantic Deep Water. An overall, speculative view of events between 20,000 and 10,000 years ago is shown in Box 1.

Other distant effects have been proposed as triggering abrupt change in the North Atlantic. For example, shifts in the frequency and amplitude of the El Niño–Southern Oscillation (ENSO) in the tropical Pacific could alter the freshwater budget of the

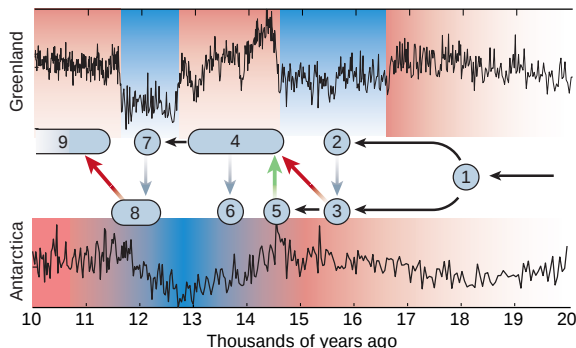
Box 1 The rough ride through deglaciation

The upper and lower panels here show high-pass filtered 'proxy' temperature records from ice cores in Greenland¹² and Antarctica¹³, respectively, indicating the abrupt coolings and warmings between 20,000 and 10,000 years ago, superimposed on the general warming trend that characterized the transition from glacial to interglacial conditions. The centre panel shows the inferred courses of events in the north and south.

(1) About 18,000 years ago, changes in Earth's orbital parameters initiate the end of the ice age through increased solar radiation. (2) Fresh water from melting northern ice sheets shuts down thermohaline circulation in the North Atlantic. (3) A 'seesaw'

mechanism (grey arrow), which connects the polar regions north and south¹⁰, enhances warming in the south. This warming turns on the thermohaline circulation in the Atlantic rapidly, by both (4) a surface advection of saline waters² and (5) a large-scale discharge of melt water in the south¹⁴, which reduces the density of deep water masses formed around Antarctica^{7,8}.

The north is now in a warm phase, and the seesaw produces (6) the cooling seen in Antarctica from around 14,000 years ago¹⁵, and also accelerates melting of northern ice sheets, which reduces the Atlantic thermohaline circulation and triggers an abrupt cooling in the north (7). This stimulates, again by the seesaw, southern warming (8). Finally, the



southern warming turns on the thermohaline circulation in the North Atlantic, leading to the final warming in the north (9) that marks the beginning of the current interglacial.

The grey vertical arrows indicate the operation of the seesaw. The

red and green arrows indicate the two different mechanisms^{2,7,8} that emerge from modelled simulations of climate, and that are proposed to be the southern cause of switching on the thermohaline circulation in the North Atlantic.

T.F.S.

tropical Atlantic, and so influence thermohaline circulation in the north of the ocean¹¹. But these mechanisms will remain speculative as long as we lack adequate temperature reconstructions or other estimates of ENSO strength during deglaciation. All in all, the tropics are probably the biggest wild card left in the game of understanding abrupt climate change.

This point takes us back to a limitation of Knorr and Lohmann's study². Because atmospheric dynamics are neglected, changes of ocean circulation patterns and their effect on atmosphere–ocean heat exchange and the freshwater balance through the hydrological cycle cannot be accounted for. So the next steps will be to perform experiments using more comprehensive models to test the importance of the processes that these authors have identified. ■

Thomas F. Stocker is in *Climate and Environmental Physics, Physics Institute, University of Bern,*

3012 Bern, Switzerland.

e-mail: stocker@climate.unibe.ch

1. Clark, P. U., Pisias, N. G., Stocker, T. F. & Weaver, A. J. *Nature* **415**, 863–869 (2002).
2. Knorr, G. & Lohmann, G. *Nature* **424**, 532–536 (2003).
3. Hogg, N. G. in *Ocean Circulation & Climate: Observing and Modelling the Global Ocean* (eds Siedler, G., Church, J. & Gould, J.) 259–270 (Academic, San Diego, 2001).
4. Broecker, W. S. *Science* **278**, 1582–1588 (1997).
5. Ganopolski, A. & Rahmstorf, S. *Nature* **409**, 153–158 (2001).
6. Wijffels, S. E. in *Ocean Circulation & Climate: Observing and Modelling the Global Ocean* (eds Siedler, G., Church, J. & Gould, J.) 475–488 (Academic, San Diego, 2001).
7. Mikolajewicz, U. *Ann. Glaciol.* **27**, 311–315 (1998).
8. Weaver, A. J., Saenko, O. A., Clark, P. U. & Mitrovica, J. X. *Science* **299**, 1709–1713 (2003).
9. Stocker, T. F., Wright, D. G. & Broecker, W. S. *Paleoceanography* **7**, 529–541 (1992).
10. Stocker, T. F. *Science* **282**, 61–62 (1998).
11. Schmittner, A., Appenzeller, C. & Stocker, T. F. *Geophys. Res. Lett.* **27**, 1163–1166 (2000).
12. Dansgaard, W. et al. *Nature* **364**, 218–220 (1993).
13. Jouzel, J. et al. *Geophys. Res. Lett.* **28**, 3199–3202 (2001).
14. Clark, P. U., Mitrovica, J. X., Milne, G. A. & Tamisiea, M. E. *Science* **295**, 2438–2441 (2002).
15. Blunier, T. et al. *Geophys. Res. Lett.* **24**, 2683–2686 (1997).

Developmental biology

How to turn inside out

Rüdiger Schmitt and Manfred Sumper

The discovery that a molecular motor of the kinesin family is involved in turning a multicellular green alga inside out might have implications for similar events in animal development.

Something quite remarkable happens to embryos of the multicellular green alga *Volvox carteri*: they turn completely inside out to establish the adult body plan. This inversion process closely resembles the initial stages of the more complex gastrulation that occurs in animal embryos — so *Volvox* could arguably be considered a simple model for analysing the principles that direct such changes in shape. As they describe in *Cell*, this idea led Nishii and colleagues¹ to re-examine and dissect the process of inversion in *Volvox*. Their concept helps to explain the biomechanics and the driving forces behind the curling of a cellular sheet.

In the embryos of most multicellular animals, sheets of cells invaginate during gastrulation, neurulation and organ formation. Gastrulation is the central process in early animal development, and occurs during the blastula stage — when the embryo consists simply of a hollow ball of cells. It involves complex movements that carry those cells whose descendants will form the future internal organs from their superficial position on the blastula to their definitive positions inside the embryo. Similarly, neurulation in vertebrates involves a complicated curling of a cellular sheet to form the neural tube, which in turn develops into the central nervous system.

Gastrulation has fascinated developmental

biologists ever since it was recognized in 1874 (for a historical review, see ref. 2). Nearly 100 years ago, the Italian embryologist Angelo Ruffini first described the appearance of elongated cells — known as bottle or flask cells — at the onset of the process in amphibians. Then, in his classic papers on amphibian gastrulation, Johannes Holtfreter^{3,4} claimed that the ability to invaginate is an innate property of flask cells. Modern investigations confirm that the number and arrangement of the flask cells are critical factors for proper initiation of invagination. But, as pointed out by Keller⁵, what flask cells do — and how, in a biomechanical sense, they do it — remains to be elucidated. *Volvox carteri*, a much simpler organism, might be the Rosetta Stone that enables researchers to unlock the problem.

Volvox is a multicellular green alga (Fig. 1) that exhibits the simplest kind of differentiation — the division of labour between just two types of cell. The adult organism consists of about 2,000 mortal somatic cells, which make up the surface of a hollow sphere, and 16 larger, potentially immortal reproductive cells just below the surface. The development of *Volvox* starts with a single reproductive cell, which undergoes a patterned sequence of 11 cell divisions. The first five divisions are symmetrical, resulting in an embryo consisting of 32 cells of similar sizes. But the sixth division of the 16 most anterior cells is

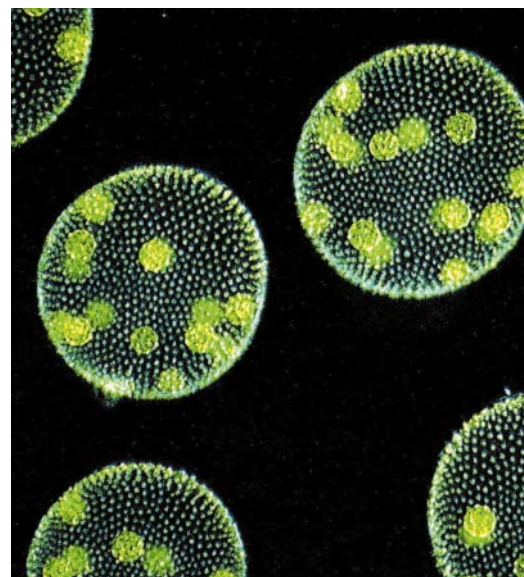


Figure 1 *Volvox carteri*. The adult organism consists of some 2,000 small somatic cells located on the surface of a hollow sphere, and around 16 large reproductive cells just below, within a transparent extracellular matrix. By a series of predetermined cell divisions each reproductive cell will become an embryo that finally turns inside out by a process called inversion, to produce a juvenile with all its cells in the adult orientation.

asymmetric, and results in the production of 16 cell pairs of unequal size. The larger cells will become the new reproductive cells.

As a result of geometrical constraints, at the end of the 11 cell divisions the embryo is inside out with respect to the adult configuration: the large reproductive cells protrude from the surface, and the bases of the developing flagella of all the somatic cells point to the interior of the hollow sphere. So cell division is followed by inversion, in which the curvature of the embryo is reversed to establish the adult configuration. Development of *Volvox* therefore resembles that of classic models of animal development, such as sea urchins and nematode worms, in that an important differentiating cell division is visibly asymmetric, and the adult configuration is attained by a gastrulation-like event.

How does inversion come about? First, the phialopore (a slit at the anterior end of the embryo) widens, and four lips of cells bend outwards and backwards over the adjacent cells — in other words, they curl outwards. The region of maximum curvature moves progressively towards the posterior pole, until the inverted region almost surrounds the posterior (non-inverted) hemisphere. At that point the posterior hemisphere 'snaps' through the opening at the equator, and the phialopore lips move to seal the gap at what is now the posterior pole.

Flask-shaped cells that are linked by a network of cytoplasmic bridges are the key

players in this dramatic shape change. Cells become flask-shaped — with a long stalk at their outer end — near the bend region, and they then move inwards relative to the network of cytoplasmic bridges, which runs through each cell and connects it to its neighbours. This is the key point in Kirk's model of inversion^{6,7}: as flask cells proceed from being linked at their widest point to being linked at their thin outermost ends, the cell sheet is forced to curl sharply outwards (Fig. 2).

Nishii *et al.*¹ now propose that this displacement of flask cells relative to the cytoplasmic bridges is driven by a motor protein — a newly discovered kinesin denoted InvA.

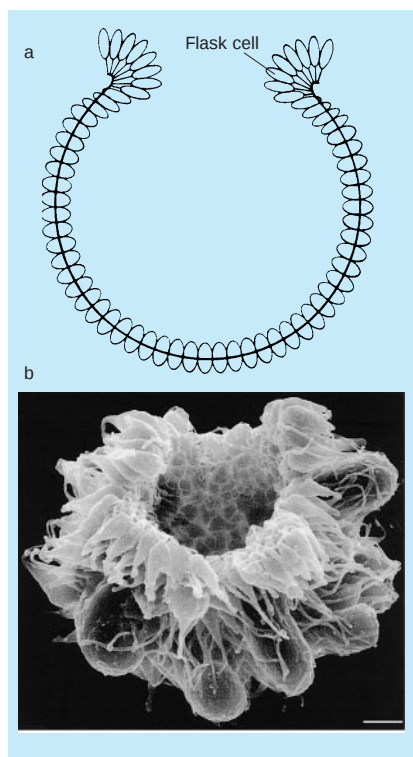


Figure 2 Inverting *Volvox*. a, Cross-section through an inverting embryo. The continuous black line represents the system of cytoplasmic bridges that connects the cells and holds them together. The force necessary for inversion is generated by a change in cell shape, together with cell movement. First the cells become flask-shaped by forming long, narrow 'stalks' at their bottom (outward) ends; then they move inwards relative to the cytoplasmic bridges to a point where they are only linked at their narrow tips. These actions produce a bend region, where the cell sheet is folded back on itself. The results of Nishii *et al.*¹ suggest that a newly discovered kinesin, InvA, is located in the cytoplasmic bridges, and that, by moving along the microtubule filaments lining the flask cells, InvA produces the force that drives the cell body past the bridges to form the bend. b, An embryo with a mutation in InvA; it is unable to invert fully. a and b are reproduced from refs 9 and 1, respectively. Scale bar, 5 μm .

The authors started by generating mutant *Volvox* embryos with defects in inversion. They then analysed one such mutant in depth, and found that although its cells became appropriately flask-shaped, they failed to move relative to the cytoplasmic bridges. Nishii *et al.* went on to show that the gene affected in this mutant encodes a kinesin, and that this molecular motor is located in the cytoplasmic bridges.

Each flask cell has cytoskeletal filaments of microtubules that run along its length, just inside the plasma membrane, and Nishii *et al.* propose that InvA molecules attempt to move downwards on these filaments. But, because the InvA molecules are anchored to the bridges, and the bridges themselves are fixed in place, InvA cannot move significantly. Instead the force produced by this motor causes the microtubules — and consequently the whole cell — to move past the cytoplasmic bridges. This perpendicular movement eventually connects the cells at their thin outermost ends and so creates a sharp bend.

The discovery of a special kinesin as part of the machinery that causes the curling of a cellular sheet in *Volvox* will stimulate a search for a related motor protein with the same function in animals. Might such a motor exist? It seems plausible: molecules such as cytoskeletal proteins are found in most

organisms, from unicellular yeasts to humans, and were probably present in the unicellular common ancestor of plants and animals. Their existence there would predetermine similar solutions to given morphogenetic problems. Moreover, if the displacement of asymmetrically shaped cells with respect to a fixed framework of cell–cell connections is the key concept underlying curling, then the cytoskeleton and cell–cell adhesion systems would have to play a key part. This would mean that similar mechanisms operate even in evolutionarily very distant⁸ organisms, from multicellular algae to animals.

Rüdiger Schmitt and Manfred Sumper are at the Institut für Biochemie, Genetik und Mikrobiologie, Universität Regensburg, 93040 Regensburg, Germany.

e-mails: rudy.schmitt@biologie.uni-regensburg.de
manfred.sumper@vkl.uni-regensburg.de

1. Nishii, I., Ogihara, S. & Kirk, D. L. *Cell* **113**, 743–753 (2003).
2. Beetschen, J.-C. *Int. J. Dev. Biol.* **45**, 771–795 (2001).
3. Holtfreter, J. *J. Exp. Zool.* **94**, 261–318 (1943).
4. Holtfreter, J. *J. Exp. Zool.* **95**, 171–212 (1944).
5. Keller, R. E. *Dev. Dyn.* **205**, 257–264 (1996).
6. Viamontes, G. I. & Kirk, D. L. *J. Cell Biol.* **75**, 719–730 (1977).
7. Viamontes, G. I., Fochtmann, L. J. & Kirk, D. L. *Cell* **17**, 537–550 (1979).
8. Kirk, D. L. in *Volvox* (eds Bard, J. B. L., Barlow, P. W., Green, P. B. & Kirk, D. L.) 16–44 (Cambridge Univ. Press, 1998).
9. Green, K. J., Viamontes, G. I. & Kirk, D. L. *J. Cell Biol.* **91**, 756–769 (1981).

Materials science

The road to diamond wafers

S. T. Lee and Y. Lifshitz

Diamond could rival silicon as the material of choice for the electronics industry, but has been held back by the difficulty of growing large enough wafers. This problem may now be solved.

Diamond is the king of gemstones. Less well known is that it could also be an outstanding semiconductor material, superior in many ways to silicon, which is currently the most widely used electronic material. Diamond devices could operate at higher temperatures (more than 400 °C) and higher power than those of silicon, as well as being faster, denser and more resistant to radiation. But practical diamond electronics will need large-area, single-crystal diamond wafers to be fabricated, analogous to the 6–12-inch silicon wafers commonly used in the semiconductor industry. Two papers from Golding and colleagues, in *Applied Physics Letters*¹ and *Diamond and Related Materials*², now show that this may be possible if sapphire wafers are used as substrates on which to grow the diamond.

Diamond can be grown on diamond ('homoeptaxy') by chemical vapour deposition: a diamond substrate, at a temperature of 600–800 °C, is exposed to an ionized

mixture of roughly 1% hydrocarbon and 99% hydrogen. Electronics-grade diamond has been made in this way³. The newly grown diamond wafer can be cut from the diamond substrate, and the process can be repeated many times, reusing the substrate. But single-crystal diamond substrates are small and expensive, so this is not a viable way to fabricate large-diameter diamond wafers.

The alternative is to grow diamond on a foreign (non-diamond) single-crystal wafer — this is 'heteroepitaxy', the oriented growth of one crystal on another. Heteroepitaxy is readily achieved if the atomic spacings in the foreign substrate match those in diamond crystals. Of the various substrates tested — such as silicon, silicon carbide, nickel, cubic boron nitride and platinum — iridium is the best found so far⁴. The quality of epitaxial growth is measured by the average angular spread of the diamond-crystal orientation along a certain direction (also called mosaicity). The minimum angle achieved is 3.9° for silicon, less than 2° for platinum, 1.5° for

silicon carbide and a mere 0.38° for iridium. The catch is that a single-crystal iridium layer has never been grown as a free-standing wafer: it requires its own single-crystal substrate. Diamond has been grown successfully on iridium that was itself grown on substrates of MgO (refs 5, 6) or SrTiO_3 (refs 4, 7, 8). But once again there is a size limitation, as MgO and SrTiO_3 wafers can be made only 2–3 inches in diameter.

The first stage of single-crystal growth is the nucleation of myriads of tiny diamond crystallites across the substrate (up to 10^{12} diamond nuclei per square centimetre). If the individual crystallites are well aligned on the substrate, the 'grain boundaries' between growing crystallites eventually merge to form a film of single-crystal diamond with relatively few defects. Schreck *et al.*^{7,8} showed that a 34- μm -thick film developed into a single-crystal diamond with a dislocation (or defect) density of about 10^9 cm^{-2} , which is comparable to that of high-quality single-crystal films of electronic materials such as gallium nitride.

But the deposition of diamond on iridium/ SrTiO_3 is plagued with problems. For instance, there is some roughening of the iridium substrate during the nucleation process, and lateral inhomogeneities often develop in the diamond film^{7,8}. And when the diamond is cooled to room temperature (from the growth temperature of 800°C), the very different thermal-expansion coefficients of iridium and diamond create thermal stress that leads to delamination of the diamond film from the substrate⁴.

Using sapphire as the substrate for iridium growth could solve some of these problems. Sapphire has many desirable properties, such as mechanical and chemical stability, and high crystalline quality. Moreover, large-diameter samples are available and relatively inexpensive. Golding and co-workers^{1,2} have taken the first step by growing high-quality iridium on sapphire wafers (Fig. 1), and they are in the early stages of studying diamond heteroepitaxy on these improved substrates. Following on from this, the optimal growth conditions and the quality of thick diamond films deposited on the new substrate should be explored. Such further experiments may well confirm sapphire as a feasible large-diameter substrate, of relatively low cost, and more stable than MgO and SrTiO_3 .

In parallel with experiments on diamond growth, researchers are also busy with the fabrication of diamond-based devices. High-performance devices have been built based on synthetic diamond crystals that were produced by high-pressure high-temperature (HPHT) techniques or by homoepitaxy of diamond on HPHT diamond crystals³; a high-quality field-effect transistor has been fabricated on heteroepitaxial diamond grown on iridium/ SrTiO_3

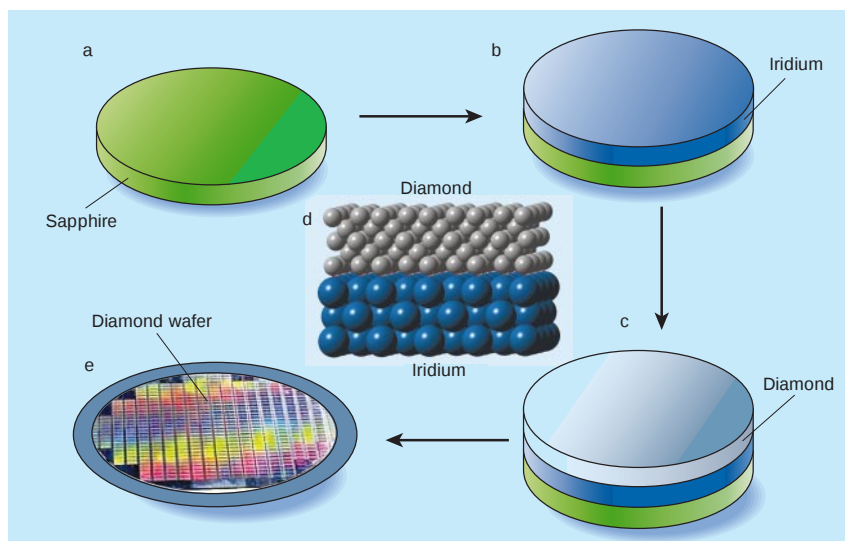


Figure 1 Heteroepitaxy for diamond wafers. Iridium is a good substrate on which to grow single-crystal diamond wafers, but an iridium layer must itself be grown on another substrate. a, Golding and co-workers^{1,2} propose a new substance for this substrate — sapphire. b, The crystal structure of sapphire directs the ordered growth of iridium on its surface, and then c, diamond is deposited on top. d, It is because the atomic spacing of the iridium crystal closely matches that of diamond that a high-quality diamond wafer can be grown. e, Just as is done currently with silicon wafers, electronic components could be cut from the large-diameter, high-quality diamond wafers made according to Golding and colleagues' prescription.

(ref. 9); and advances are being made in diamond doping, in which the electronic properties of the material are manipulated by adding small amounts of impurity atoms¹⁰.

An attractive, rival approach has been proposed that uses silicon as the substrate for iridium growth, with a CaF_2 interlayer¹¹, to take advantage of the ready availability of large, low-cost silicon wafers. Nevertheless, the development of diamond–iridium–sapphire heteroepitaxy by Golding and co-workers^{1,2} and the progress made in device fabrication raise unprecedented hope for the future of large-scale diamond electronics. ■

S. T. Lee and Y. Lifshitz are at the Center of Super-Diamond and Advanced Films, and in the

Department of Physics and Materials Science, City University of Hong Kong, Hong Kong SAR, China. e-mail: apannale@cityu.edu.hk

1. Dai, Z., Bednarski-Meinke, C., Loloee, R. & Golding, B. *Appl. Phys. Lett.* **82**, 3847–3849 (2003).
2. Bednarski, C., Dai, Z., Li, A.-P. & Golding, B. *Diamond Related Mater.* **12**, 241–245 (2003).
3. Isberg, J. *et al. Science* **297**, 1670–1672 (2002).
4. Schreck, M., Roll, H. & Stritzker, B. *Appl. Phys. Lett.* **74**, 650–652 (1999).
5. Ohtsuka, K., Fukuda, H., Suzuki, K. & Sawabe, A. *Jpn. J. Appl. Phys. Pr* **2** **36**, L1214–L1216 (1997).
6. Tsubota, T. *et al. Diamond Related Mater.* **9**, 1380–1387 (2000).
7. Schreck, M. *et al. Appl. Phys. Lett.* **78**, 192–194 (2001).
8. Schreck, M. *et al. Diamond Related Mater.* **12**, 262–267 (2003).
9. Kubovic, M. *et al. Diamond Related Mater.* **12**, 403–407 (2003).
10. Teukam, Z. *et al. Nature Mater.* **2**, 482–486 (2003).
11. Lee, C. H., Qi, J., Lee, S. T. & Hung, L. S. *Diamond Related Mater.* **12**, 1335–1339 (2003).

Theoretical biology

Safeguards and spurs

Stephen C. Stearns

Biological traits are buffered against genetic and environmental upset by a process called canalization. New work suggests this may be a general feature of regulatory gene networks, selected to survive gene loss.

One of the biggest challenges facing twenty-first-century biologists is understanding how genes are connected to the traits seen in whole organisms. At a basic level, we know that the overall appearance and behaviour of an organism are determined by the interaction between its genes and the environment. Yet this interaction is incredibly complicated.

For instance, some kind of developmental buffering seems to ensure that similar traits are produced despite genetic changes (mutations) and environmental perturbation. A better understanding of this buffering effect, known as canalization, could change how we think about evolution and development.

Theoretical insights^{1–3} and experimental

Neurobiology

Motor proteins branch out

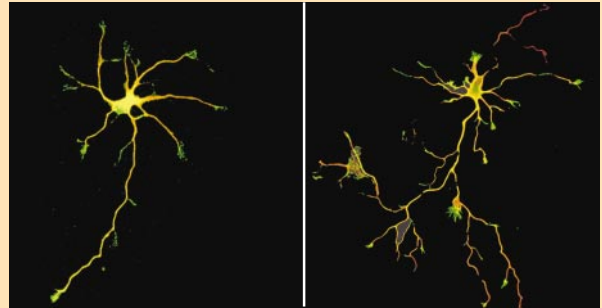
Molecular motors are perhaps best known for their ability to transport cargo around inside cells. To do so, these tiny vehicles move along defined intracellular 'tracks'; motor proteins of the kinesin family, for instance, move on microtubule filaments. But some kinesins have a quite separate function, at least in cultured cells: they can break down microtubules into their constituent parts. Writing in *Cell* (114, 229–239; 2003), Noriko Homma *et al.* show that the kinesin KIF2A has this activity *in vitro* — and they propose that this function is essential *in vivo* for the brain to develop normally.

The authors started by generating mice that lack KIF2A, and found that the animals died within a day of birth with severe brain abnormalities. Delving deeper, Homma *et al.* discovered defects in the collateral branches of neurons. During development, nerve cells extend projections — axons — to form appropriate connections in the

brain. Sometimes, branches also extend outwards from the axons to form further connections. These collateral branches generally remain short until the primary axon has found its way.

But that doesn't appear to happen in KIF2A-deficient mice. The authors took neurons from the hippocampus of normal and mutant mice and cultured the cells for two days. The normal neurons (left-hand image) extended a single primary axon, with some short collateral branches. The mutant nerves, by contrast, developed abnormally long collateral branches, which rebranched several times (right-hand image).

Video analysis showed that, in wild-type neurons, any collateral branches that formed actively shrank back. But in the mutant cells the branches continued growing. So KIF2A seems to be necessary to stop the branches lengthening. How does it do this? Homma *et al.* found



that the protein can depolymerize microtubules within extending neuronal protrusions, and that this activity is reduced in KIF2A-deficient cells. Moreover, when the authors studied the polymerization and depolymerization of microtubules in another type of cultured brain cell (it's difficult to do this in neurons), they found that, if KIF2A was missing, microtubules carried on growing when they reached the edge of the cell. But in the presence of KIF2A, microtubules either stopped growing or showed

cycles of growth and shrinkage.

It seems, then, that microtubule growth drives the protrusion of collateral branches — possibly by pushing out the edge of the cell — and that KIF2A normally keeps this process under control by depolymerizing microtubules when they reach the edge. It remains possible that KIF2A has an indirect role: perhaps it carries microtubule-depolymerizing molecules to the edge. But the protein's *in vitro* activity would seem to argue against this. **Amanda Tromans**

results^{4,5} alike are enlivening discussion of canalization. Theoretically, it seems that canalization can be a by-product of the existence of genetic regulatory networks that maintain their function when any given gene is knocked out³. On the experimental side, the protein and molecular chaperone Hsp90 has properties that suggest it could be a central player in a general canalizing mechanism^{4,5}. On page 549 of this issue, Bergman and Siegal⁶ advance the discussion by showing — with both a simulation model and new analysis of data on the effects of gene knockouts in yeast — that most genes in regulatory networks have canalizing properties.

The authors first set up computer simulations of gene networks, in which each virtual gene regulates the expression of itself and every other gene. A matrix showing the regulatory relationships is the 'genotype' — the representation in the model of the set of genes relevant to the issue at hand. The traits that might be buffered (the 'phenotype') are represented by the equilibrium profile of gene expression that results from the regulatory relationships. The starting point is arbitrary, and not yet at equilibrium. A gene is then knocked out at random by setting its value in the matrix to zero, and the population is allowed to evolve.

The results were striking: the knockout populations showed, in general, much more phenotypic variation between individuals

than did populations in which genes were not deleted. And simulated populations with a supply of knockout mutations evolved more rapidly to an equilibrium (optimum) phenotype than did simulated populations without such a supply. The authors propose that the knocked-out regulatory genes usually have canalizing effects on phenotypic variation that are lost when the genes are knocked out — releasing previously hidden variation. In my view, however, this in itself does not provide evidence of canalization, for an increased supply of mutations with diverse effects would be expected to accelerate evolution whether or not a regulatory network with canalizing effects existed.

On the other hand, Bergman and Siegal also analysed experimental data from a large-scale project with the yeast *Saccharomyces cerevisiae*; this project aims to delete each yeast gene in turn, and to study the effects. The genetic background of the yeasts is identical, save for the gene knockout, so here the authors were studying the effects of environmental variation on phenotype. They found that deleting a gene at random increased environmentally induced variation in the expression of other genes — again suggesting that such phenotypic variation is usually canalized.

Bergman and Siegal interpret these findings as providing support for the notion of 'evolutionary capacitance'. By this they mean

that evolution is accelerated when genetic or environmental stress disturbs canalizing mechanisms. The disturbance reveals genetic and phenotypic variation that had previously been hidden, and which now becomes available to fuel an evolutionary response that would not have been possible had the canalizing mechanism not been functionally compromised.

How do these results fit with previous work? From experiments on fruitflies and thale cress^{4,5} we know that trait variation increases when Hsp90 is functionally compromised, and that the function of Hsp90 can be compromised by environmental stress. From theory we know that knockout mutations in genetic regulatory networks reveal hidden genetic and phenotypic variation, and speed the approach to a new optimum. And Bergman and Siegal's analysis of yeast knockouts suggests that knocking a gene out at random in a genetically uniform background increases environmentally induced variation in the expression of other genes.

So far so good. But a crucial element in the theory is missing from the data: we do not yet have an experimental demonstration that compromising a canalizing mechanism accelerates evolution to a new optimal phenotype. We only know that the amount of genetic variation that is expressed as phenotypic variation is increased. That such

experiments are hard to do is only a temporary excuse; it does not scratch the logical itch. Bergman and Siegal's work makes it more plausible that capacitance effects will be found. Yet there are at least three reasons why their results are not yet sufficient to convince a sceptic that capacitance is real.

First, in their simulations the authors looked only at gene expression (the 'transcriptome'), rather than at anything approaching the complexity of whole-organism phenotype. So their results leave us uncertain about what might happen in between. Variation in gene expression might in fact lead to canalization at the whole-organism level, if appropriately regulated transcriptome variation is part of the mechanism that canalizes whole-organism traits.

Second, variation released in the transcriptome could also produce changes in reproduction and mortality rates, which determine the strength of natural selection. If increased genetic variation were accompanied by reduced variation in reproductive success, or by reductions in the correlation between variation in traits and variation in reproductive success, the rate of evolution would not necessarily change. In fact, if the effects were strong enough, the rate of evolution might even decrease.

Third, Bergman and Siegal's analysis of yeast data reveals an increase in phenotypic, not genetic, variation when genes are knocked out at random. Yet capacitance requires that the increased variation be, at least in part, genetic. It has been asserted^{7,8} that increased phenotypic variation implies increased genetic variation in such circumstances. But it might not.

So the existence and importance of evolutionary capacitance remain uncertain. But if canalization does turn out to be a pervasive property of genetic networks — one that evolves rapidly and serves as an evolutionary capacitor — a key feature of the genotype–phenotype connection will have been clarified. And that would stimulate further questions. One is how part of the organism's response to environmental change remains canalized while others become 'plastic', or flexible. Also, how do the genetic networks that might be producing canalization as a by-product interact with the mechanisms that determine plasticity (if they interact at all)? Does evolutionary capacitance explain the periods of stasis and rapid change that are seen in experimental studies of evolution — or perhaps even in the fossil record? The answers to such questions may seem beyond reach, but in fact we can start right now with evolutionary experiments on the functional genomics of microorganisms. ■

Stephen C. Stearns is in the Department of Ecology and Evolutionary Biology, Yale University, New Haven, Connecticut 06520-8106, USA.
e-mail: stephen.stearns@yale.edu

1. Wagner, A. *Evolution* **50**, 1008–1023 (1996).
2. Wagner, G., Booth, G. & Baghieri-Chiachian, H. A. *Evolution* **51**, 329–347 (1997).
3. Siegal, M. L. & Bergman, A. *Proc. Natl Acad. Sci. USA* **99**, 10528–10532 (2002).
4. Rutherford, R. L. & Lindquist, S. *Nature* **396**, 336–342 (1998).
5. Quetsch, C., Sangster, T. A. & Lindquist, S. *Nature* **417**, 618–624 (2002).
6. Bergman, A. & Siegal, M. L. *Nature* **424**, 549–552 (2003).
7. Stearns, S. C., Kaiser, M. & Kawecki, T. J. *J. Evol. Biol.* **8**, 539–557 (1995).
8. Meiklejohn, C. D. & Hartl, D. L. *Trends Ecol. Evol.* **17**, 468–473 (2002).

on the atoms. In other words, the conduction electrons have a slight preference energetically to have their spins oriented in the opposite direction to that of the magnetic moments localized on the atomic sites in the lattice. As the temperature falls below the value that characterizes this ordering preference, the local atomic moments become progressively screened and the entropy associated with the local magnetic degeneracy 'dresses' the conduction electrons, dramatically increasing their effective masses and making them into 'heavy fermions'. This reduction of atomic moments drives down the magnetic ordering temperature, and as the local interaction between magnetic moment and conduction electron becomes stronger, the moment and the magnetic ordering temperature simultaneously drop to zero — this is the quantum critical point.

Much of the interest in quantum critical physics stems from its relevance to the physics of phenomena that involve correlated behaviour of electrons — the prime example here being heavy-fermion superconductivity, the resistance-less flow of current at very low temperatures. Discovered³ in 1979, this particular variant of superconductivity demanded a new understanding of the nature of heavy-fermion materials, as their magnetic character had been thought to be hostile to the occurrence of superconductivity. In the 1990s, it became clear that the systematic appearance of superconductivity in some heavy-fermion materials happens in the vicinity of what has since been recognized as a quantum critical point⁴. Although this is not true for all heavy-fermion materials (including YbRh₂Si₂, studied by Custers *et al.*²), the tantalizing link between superconductivity and quantum critical points has driven efforts to understand both phenomena.

Classically, critical phenomena are narrowly confined to the region around a critical point, but, in contrast, the influence of a quantum critical point spreads through the phase diagram, well away from the point itself. The effects are many and can be parametrized through properties that vary with a power of temperature, such as electrical resistivity, magnetic susceptibility and specific heat. In their experiments, Custers *et al.*² have traced the variation of such properties as they tuned their heavy-fermion material, YbRh₂Si₂, into a quantum critical point by applying magnetic fields.

In these experiments, to probe the quantum critical point more deeply, the authors 'doped' their samples of YbRh₂Si₂, replacing some of the silicon atoms with germanium. Germanium atoms are slightly larger than silicon ones, and so increase the pressure in the atomic lattice and drive down the strength of the magnetic field at the critical point to almost zero. Custers *et al.* have now provided the most complete study so far of the heavy-fermion quantum critical point, combining

Condensed-matter physics

Singular behaviour

Zachary Fisk

Quantum fluctuations at absolute zero may push a system into a different phase or state. The 'quantum critical point' at which this happens in certain materials has now been probed in greater detail.

More than a century of probing the physics of magnetism has only served to heighten its interest for physicists. Over the past decade, a new focus has evolved — the behaviour of magnetic materials at a 'quantum critical point'. Critical points are perhaps more familiarly associated with phase transitions such as those between water, steam and ice. But at zero temperature there are quantum phase transitions, fluctuations between different states of a phase diagram that are governed by Heisenberg's uncertainty principle. In certain magnetic materials, the temperature at which the system takes on a magnetically ordered state can be tuned, either chemically

or through some external influence, towards absolute zero, the transition point being a quantum critical point¹. The question is, at a quantum critical point, do we see something new that does not occur in classical systems? On page 524 of this issue, Custers *et al.*² go some way towards an answer, through the closest approach yet to a quantum critical point in a 'heavy-fermion' material.

Heavy-fermion materials, a class of so-called correlated-electron materials, are metallic compounds with a chemically ordered lattice of magnetic ions, in which conduction electrons interact antiferromagnetically with the local magnetic moments

chemical doping and magnetic-field measurements to very low temperatures. Chemical doping, however, introduces inhomogeneities in the material that might easily confuse the interpretation of experiments. But in this case the results for the germanium-doped material overlap convincingly with those of the pure compound. This overlap is by no means a foregone conclusion: magnetic field and chemical (or here, size) doping are not in general interchangeable. But the singular nature of the quantum critical point is revealed in these measurements: the mass of the 'dressed' electrons diverges towards infinity.

The temperature dependence of physical properties, in particular how these quantities scale with an exponent of temperature, is a useful diagnostic close to the quantum critical point. This scaling behaviour — of specific heat, electrical resistivity and magnetic susceptibility — is used in attempts to model quantum critical behaviour, and to describe the physics of the fluctuations. Custers *et al.*² have uncovered a linear temperature-dependence of electrical resistivity over a remarkably large range, from 10 mK to 10 K (as shown in Fig. 1 on page 524). Moreover, the new data reveal that, as the temperature changes and the system moves away from the quantum critical point, temperature is the only energy scale necessary to describe the evolution. Similarly, when the magnetic field

is varied, that quantity sets the energy scale for variations in the material's properties.

There is obvious danger in drawing conclusions from such power-law fits, which can only be suggestive, but our modern understanding requires that if there is a critical point, there should be scaling. Theoretical work^{5,6} has shown that it is reasonable to think of quantum phase transitions as classical phase transitions in higher dimensions, but the question remains whether the highly developed scaling concepts of classical critical phenomena can be simply borrowed here.

The measurements made by Custers *et al.*² show just how strongly the phase diagrams of heavy-fermion materials are organized by their singularities, the quantum critical points. The goal is to understand the nature of this organization at a deeper level; the dream is that some essentially new ground state may be found in the quantum-critical-point regime. ■

Zachary Fisk is in the Department of Physics, Florida State University, Tallahassee, Florida 32306, and the Los Alamos National Laboratory, K774, Los Alamos, New Mexico 87545, USA.

e-mail: fisk@magnet.fsu.edu

1. Sachdev, S. *Science* **288**, 475–480 (2000).
2. Custers, J. *et al.* *Nature* **424**, 524–527 (2003).
3. Steglich, F. *et al.* *Phys. Rev. Lett.* **43**, 1892–1896 (1979).
4. Mathur, N. D. *et al.* *Nature* **394**, 39–43 (1998).
5. Hertz, J. A. *Phys. Rev. B* **14**, 1165–1184 (1976).
6. Millis, A. J. *Phys. Rev. B* **48**, 7183–7196 (1993).

Cell biology

Moving inside membranes

Katsuyoshi Mihara

The mechanism that inserts proteins into the membranes of cellular organelles was thought to be well understood. But studies in yeast reveal that this process is sometimes more complicated than had been suspected.

Rather like the organs of the human body, the 'organelles' of plant, animal and yeast cells are specialized compartments that fulfil specific functions. Each organelle is bounded by a lipid membrane, which contains 'translocase complexes' that ferry proteins from outside the compartment to the inside. But organelle membranes are more than just barriers; besides the translocase complexes, they contain many other proteins, which often adopt intricate configurations within the membrane itself. The prevailing view is that these proteins are sorted and assembled in the membrane by the same translocase complex that transports proteins across the membrane to the organelle interior. But on page 565 of this issue, Wiedemann *et al.*¹ report an unexpected finding from their studies in yeast mitochondria. They show that additional machinery, besides the translocase complex, is required to sort and assemble mitochondrial outer-membrane

proteins that have a complicated conformation — including the translocase proteins themselves.

Mitochondria, the powerhouses of the cell, are bounded by not one, but two membranes. Some mitochondrial proteins reside in one of these membranes, some occur in the space between the membranes, and yet others are at the heart of the mitochondrion (Fig. 1, overleaf). All of these proteins are synthesized as precursor proteins (preproteins) inside the cell and are shuttled across the membranes by TOM and TIM complexes — preprotein translocases of the outer and inner membranes, respectively². The TOM complex forms a channel in the outer membrane, called the general insertion pore, through which nearly all mitochondrial preproteins pass. The channel is made by the protein Tom40. This protein chain spans the membrane many times, forming an intricate pore-shaped structure. The entire TOM complex, however, is composed of many



100 YEARS AGO

The additions to the Zoological Society's Gardens during the past week include a Sooty Mangabey (*Cercocebus fuliginosus*) from West Africa, presented by Mrs. Watkins; a Ring-tailed Lemur (*Lemur catta*) from Madagascar, presented by Mr. H. P. Jacques; a Suricate (*Suricata tetradactyla*) from South Africa, presented by Captain C. P. Harvey; two Kinkajous (*Cercoptes caudivulvulus*) from South America, presented by Miss C. Wallace Dunlop; a Himalayan Whistling Thrush (*Myiophonus temmincki*), a Blue-winged Siva (*Siva cyanouroptera*), a Lesser Blue-winged Pitta (*Pitta cyanoptera*) from the Himalayas, presented by Mr. E. W. Harper; ... two Wanderoo Monkeys (*Macacus silenus*) from Malabar, a Common Crowned Pigeon (*Goura coronata*), a Sclater's Crowned Pigeon (*Goura sclateri*) from New Guinea... two Indian Rollers (*Coracias indica*), three Pond Herons (*Ardeola grayi*), five Scarlet-backed Flower-peckers (*Dicaeum cruentatum*), two Two-banded Monitors (*Varanus salvator*) from India, deposited. From *Nature* 30 July 1903.

50 YEARS AGO

Traité de zoologie Anatomie, systématique, biologie. Publié sous la direction de Prof. Pierre-P. Grassé. Tome 1, Fascicule 1: Phylogénie; protozoaires, généralités; flagellés. (Paris: Masson et Cie., 1952.) 9,600 francs. This is another volume of the now well-known "Traité de Zoologie", issued under the editorship of Prof. Pierre-P. Grassé, of the Sorbonne. While the seventh to appear, it is in fact the first fascicle of the first volume of the series and is in every way worthy of its predecessors. Its 1,071 pages are provided with 829 illustrations (there is no Fig. 285), 694 in line-drawings, 116 in half-tone from wash drawings, 18 in half-tone of photographs and 1 in line and colour... The present volume deals only with the general introductory matters and the sub-phylum Flagellata. The introductory accounts of the structure, physiology, nuclear behaviour, life-cycle and biology of the various categories form a valuable part of the work. Under the systematics of each class there is, wherever possible, a section dealing with its fossil members, a reminder of the present interest in microfossils. It is regrettable that the exigencies of the times are reflected in the price, for this highly commendable book should be readily accessible to every zoologist. From *Nature* 1 August 1953.

different Tom proteins, which together have a relative molecular mass of 400,000 (400K).

But how are the mature TOM complexes themselves assembled and inserted into the membrane? Some details were already known. Newly synthesized Tom40 precursor protein is recognized by the import receptor Tom20 and is then shuttled across the outer membrane through pre-existing TOM channels. The assembly of new translocases then occurs through two intermediate structures³. First, the pore-forming Tom40 preprotein assembles with Tom5 to form a 250K intermediate complex (intermediate I), which is associated with the inner surface of the outer membrane. Second, this complex rearranges into a 100K structure (intermediate II), concomitant with the integration of Tom40 into the membrane. This step is followed by the sequential addition of several other Tom proteins to form the mature TOM complex³.

In the new study, Wiedemann *et al.*¹ asked whether a protein called Mas37 was involved in the assembly of TOM complexes. This protein was first identified during the screening of yeast mutants defective in phospholipid metabolism, but mutation of the Mas37 gene has many other adverse effects on mitochondrial function⁴. Although Mas37 is not part of the mature 400K translocase, it is found on the outer mitochondrial membrane and has been proposed to act as a preprotein receptor. But its precise involvement in the assembly of TOM has been unclear⁵. To investigate this, Wiedemann *et al.* tracked the Tom40 preprotein after import into yeast mitochondria lacking Mas37. They found that the formation of TOM intermediates I and II, as well as mature TOM, was blocked in these cells. Strikingly, however, the import and assembly of preproteins destined for inner mitochondrial sub-compartments were unaffected.

Wiedemann *et al.* next examined whether Mas37 was involved in the assembly of other outer-membrane proteins besides Tom40. They found that assembly of the proteins porin and Mdm10 — which have complicated structures that span the outer membrane many times — was inhibited in Mas37-deficient yeast cells, but that simpler outer-membrane proteins were assembled as normal. These results suggest that Mas37 only participates in the assembly and integration of outer-membrane proteins that have complicated conformations.

Could Mas37 be a component of the TOM assembly intermediates I or II? The authors found that anti-Mas37 antibodies bound to intermediate I, but not to intermediate II or mature TOM, suggesting that Mas37 is indeed an integral part of the 250K intermediate I complex. In support of this finding, Tom40 accumulated within a structure of around 210K in the absence of Mas37 (which, as its name suggests, has a relative

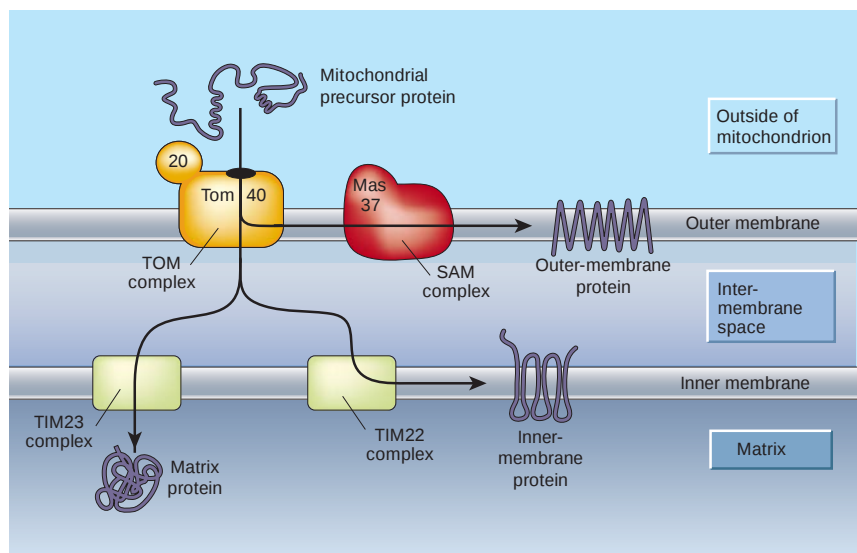


Figure 1 Sorting and assembly pathways of mitochondrial precursor proteins. All mitochondrial preproteins are recognized by import receptors (Tom20 is the major import receptor) and are then imported through the translocase of the outer mitochondrial membrane (TOM complex). Preproteins destined for the inside of the mitochondrion (the matrix) are then shuttled through the inner membrane by the TIM23 complex, whereas those destined for the inner membrane are transferred across the intermembrane space to the TIM22 complex, through which they are inserted into the membrane. Wiedemann *et al.*¹ have found that outer-membrane proteins with a complicated topology pass through the TOM complex, then become integrated in the membrane with the assistance of a separate sorting and assembly complex (SAM).

molecular mass of 37K). Furthermore, in the absence of Tom40, Mas37 was found in a 210K complex that was distinct from the mature TOM complex. The authors conclude that the 210K structure containing Mas37 is a complex that makes up the bulk of intermediate I, and that also recruits certain proteins destined for the outer membrane, such as Tom40. They call this complex SAM, because it appears to be necessary for the correct sorting and assembly of outer-membrane proteins.

These findings suggest that outer-membrane preproteins are imported through TOM, then recognized and assembled by SAM. In the case of Tom40, when SAM recognizes and assembles this protein, the resulting complex makes up intermediate I. To confirm that Tom40 was initially imported through the TOM channel, Wiedemann *et al.* saturated pre-existing TOM channels with a preprotein destined for the inside of the mitochondrion, and then tried to import Tom40 preprotein. They found that the assembly of the SAM–Tom40 complex (intermediate I) was blocked, indicating that Tom40 does need to be transported through the TOM channel before it is handed over to SAM.

Wiedemann and colleagues' results change our views on the sorting and assembly of proteins in the organelle membrane. It had generally been assumed that the sorting of membrane proteins always occurred within the translocation machinery itself. Wiedemann *et al.* show that in the mitochondrial

outer membrane this is not the case. Instead, although all mitochondrial precursor proteins initially pass through the translocation channel, some proteins destined for the outer membrane are subsequently recruited by the SAM complex, which helps them to integrate into the membrane and attain their correct conformation (Fig. 1).

So far, this finding is restricted to mitochondrial outer-membrane proteins that have complicated conformations. It remains to be seen whether the protein translocases of other organelles^{6,7} also possess distinct sorting and assembly machinery for inserting complex proteins into the membrane. Perhaps there are unique mechanisms operating in the membranes of other organelles that are as yet unidentified. How the TOM or SAM complex discriminates between preproteins for the outer membrane and those for other mitochondrial subcompartments is another unanswered question. Dissecting the molecular cues that underlie this process should be the next goal.

Katsuyoshi Mihara is in the Department of Molecular Biology, Graduate School of Medical Science, Kyushu University, 812-8582 Fukuoka, Japan.
e-mail: mihara@cell.med.kyushu-u.ac.jp

1. Wiedemann, V. *et al.* *Nature* **424**, 565–571 (2003).
2. Pfanner, N. & Geissler, A. *Mol. Cell Biol.* **2**, 339–349 (2001).
3. Model, C. *et al.* *Nature Struct. Biol.* **8**, 361–370 (2001).
4. Gratzner, S. *et al.* *J. Cell Biol.* **129**, 25–34 (1995).
5. Ryan, M. T., Müller, H. & Pfanner, N. *J. Biol. Chem.* **274**, 20619–20627 (1999).
6. Johnson, A. E. & van Waas, M. A. *Annu. Rev. Cell Dev. Biol.* **15**, 799–842 (1999).
7. Jarvis, P. & Soll, J. *Biochim. Biophys. Acta* **1541**, 64–79 (2001).

Neurobiology

Whiskered to perfection by serotonin

Neuron 39, 343–352 (2003)

Rodents can explore their world via their whiskers — specialized hairs that vibrate rhythmically and feed sensory information to the brain. A group of cells found in the brain stem is thought to act as a pacemaker that stimulates specific facial motor neurons to generate whisker movements. Alexis Hattox *et al.* now show that pacemaker cells use the neurotransmitter serotonin (5-HT) to communicate with the 'whisking' motor neurons.

The authors found that when whisking motor neurons were treated with 5-HT *in vitro*, they showed rhythmic, repetitive bursts of activity. And when the authors stimulated serotonin-releasing neurons that are thought to be part of the pacemaker *in vivo*, rats moved their whiskers. The neurotransmitter appears to bind to and activate specialized subsets of 5-HT receptor on the whisking motor neurons: cells treated with compounds that block the 5-HT₂ or 5-HT₃ receptor types were unable to fire. Moreover, rats treated with 5-HT inhibitors, and allowed to roam around, seemed to have lost their natural whisker rhythm.

Hattox *et al.* conclude that a group of serotonin-releasing cells in the brain stem regulates rhythmic whisking. These cells may in turn receive input from the cerebral cortex, forming a complex neuronal circuit.

Helen R. Pilcher

Cosmology

Support for dark energy

Preprint at <<http://arxiv.org/astro-ph/0307335>> (2003)

The existence of dark energy — the origin of the repulsion causing the expansion of the Universe to accelerate — has so far hinged on the observation of dimmer-than-expected distant supernovae. But Ryan Scranton *et al.* have now found independent and more direct evidence. They show that the cosmic microwave background (CMB) radiation is slightly 'hotter' where there are more galaxies.

This 'warming' arises because CMB photons have their frequencies altered by gravitational fields. When these photons pass by a concentration of mass (such as a galaxy) and fall into its gravitational well, they gain energy. But then they lose the same amount of energy as they climb out of the well. Or at least, they would do if all they encountered was normal matter. Because dark energy is gravitationally repulsive, it causes a gravitational well to become shallower as a photon passes through it. The photon then exits with a slight energy gain.

This would make the CMB hotter where there are more galaxies.

Scranton *et al.* have now identified just such a positive correlation between the CMB temperature map and the distribution of about 25 million galaxies. Assuming that (as is generally thought) the Universe is flat, dark energy seems to be the only explanation for the correlation.

Philip Ball

Biomaterials

Take a break

Development 130, 4123–4133 (2003)

Embryonic bone development and adult fracture repair may be two sides of the same coin. Both processes involve progenitor cells, which clump together and then produce cartilage and bone. And both require a blood supply to be established, to keep the cells alive.

Céline Colnot *et al.* now report another similarity. The group looked at how adult bones repair themselves in the absence of matrix metalloproteinase 9 (MMP9), a molecule known to regulate blood-vessel formation in developing bones. They found that fractures in adult mice lacking the *Mmp9* gene took longer to heal than normal, partly because blood vessels took longer to form in the damaged area. The situation could be rectified by injecting vascular endothelial growth factor (VEGF) into the fracture site, where it stimulated the recruitment and/or production of blood-vessel cells and cartilage- and bone-making cells. This suggests that VEGF and MMP9 may act together to coordinate skeletal repair.

Colnot *et al.* speculate that adult bone repair and early bone development share similar molecular mechanisms that rely on MMP9. As a result, embryonic programmes may be reactivated in later life to help fix broken bones.

Helen R. Pilcher

Animal behaviour

Bowerbird blues

Anim. Behav. 65, 1077–1083 (2003)

Male strategies to entice a female to mate may, according to one hypothesis, tap into a pre-existing bias in the female's sensory system. Joah Madden and Kate Tanner have tested this possibility in five species of bowerbird, the males of which attract females by building and decorating elaborate displays, or bowers. Among the decorations they use are fruits of different colours, employed as ornaments not food. Madden and Tanner searched for such a pre-existing bias, sculpted by foraging preferences, by presenting males and females of several species, held in Australian zoos, with grapes artificially coloured in different hues.

The five species did not show a common bias in sensory preference for any single



A female satin bowerbird.

colour. The sample size precluded a full statistical analysis, but suggested that the birds fell into two groups, with closer relatives preferring similar colours. Stronger support for the sensory-bias hypothesis came, however, from two intensively studied species, the satin and regent bowerbirds. In these cases Madden and Tanner did indeed find a relationship between male preferences for fruit colour used in bower decoration (blue) and female preferences for the colour of fruit eaten.

Tom Clarke

Microbiology

Aspirin bites at bugs

J. Clin. Invest. 112, 222–233 (2003)

The all-purpose drug aspirin helps to reduce the virulence of the common hospital bug *Staphylococcus aureus*. Leon Iri Kupferwasser *et al.* show that it does so by directly affecting the bacterium, not the patient.

The authors soaked *S. aureus* in the main *in vivo* metabolite of aspirin, salicylic acid, and injected the bugs into rabbits. The heart-valve abscesses that formed were roughly half the size of those produced on injecting untreated bacteria. This suggests that aspirin acts on the bacteria, contradicting the previous assumption that it fights infection by stopping blood platelets from clumping together in infected areas of the body.

Kupferwasser *et al.* also found that salicylic acid switches on the *sigB* system in *S. aureus* that enables the bacterium to respond to stressful conditions. This triggers a drop in the production of fibronectin-binding protein — used to bind to and infect host tissues — and the toxin α -haemolysin. Finally, bugs lacking the *sigB* gene no longer reacted to salicylic acid.

The authors suggest that patients with *S. aureus* infections, which are common following the use of intravenous tubes and prosthetic implants, might benefit from treatment with aspirin alongside standard antibiotics. Its use might also help to avoid spiralling antibiotic resistance in the bacterium.

Helen Pearson

Froghopper insects leap to new heights

An innovative leaping action propels these bugs to the top of the insect athletic league.

There are two basic body designs for jumping that enable many animals to escape from predators, to increase their speed of locomotion or to launch into flight¹. Animals with long legs (bush babies, kangaroos and frogs, for example) have a levering power that enables them to use less force to jump the same distance as short-legged animals of comparable mass, whereas those with short legs must rely on the release of stored energy in a rapid catapult action. Insects exploit both designs: bush crickets use the leverage provided by long legs², fleas use stored energy to power their short legs³, and grasshoppers⁴ combine features of each. Fleas are considered to be the champion jumpers, but here I show that froghoppers (spittle bugs) are in fact the real champions and that they achieve their supremacy by using a novel catapult mechanism for jumping.

Froghoppers, *Philaenus spumarius* (Linnaeus), have a mass (m) of 12.3 ± 0.7 mg (mean $\pm$ s.e.m., $n = 11$) and are 6.1 ± 0.2 mm long. In preparing to jump (in ten jumps by four adults), the front and middle legs raise the front of the body and the thrust is then provided by the simultaneous extension of both hind legs in less than 1 ms (Fig. 1). These lift the body at a take-off angle of $58 \pm 2.6^\circ$ and the mean take-off velocity (v) is 2.8 ± 0.1 m s⁻¹ (or, for the best jumps, 4 m s⁻¹). The body is thus accelerated at $2,800\text{--}4,000$ m s⁻², which is equivalent in the best jumps to 408g.

The energy ($0.5mv^2$) required for an average jump is 49 microjoules, with the best jumps needing 100 μ J. This corresponds to an average power output of 100 mW if the acceleration (a) is conservatively applied in 1 ms. About 11% ($n = 5$) of the body weight is attributable to the pair of trochanteral depressor muscles in the body that generate the rapid movements of the hind legs, giving a power output of 36 W g⁻¹ and a maximum for the best jumps of 74 W g⁻¹.

This is much greater than the power output from the jumping muscles in a locust and about the same as that in a flea, but less than that in a click beetle^{5,6}. The force (mf) exerted in an average jump is 34 mN. For the highest jumps, the force increases to 49 mN and, because the cross-sectional area of this muscle is 1.2 mm², it equates to 41 kN m⁻² of muscle⁷. The average height achieved is 428 ± 26.1 mm, with the highest jumps reaching 700 mm.

The forces powering the jump could not be produced by direct muscle contractions over the short distances and brief time available, indicating that muscular force

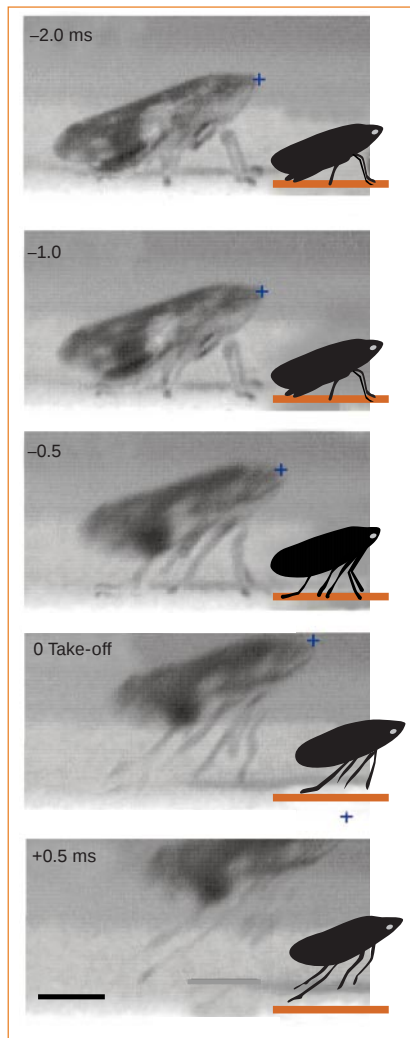


Figure 1 Sequential images of a spontaneous jump by *Philaenus spumarius* captured at 2,000 frames per second and an exposure time of 0.25 ms with a high-speed camera (Redlake Imaging, San Diego) and associated computer. Take-off at time 0 ms occurs within two frames, or less than 1 ms, of the first movement of the hind legs between frames -1.0 and -0.5 ms. The crosshairs mark the position of the front of the head and the tracings on the right show the movements of the body and legs. Scale bar (bottom panel), 2 mm.

must be generated by a slow contraction in advance of the movement, storing energy which is then released rapidly. The hind legs are roughly half the length of the body and 1.5 times that of the other legs. Before take-off, they are rotated forwards at the coxo-trochanteral joints so that the femora are tucked between the thorax and the middle legs. The tibiae are also flexed about the femora so that they are parallel with the long axis of the body. A ridge on the dorsal femur rides over and engages in front of a

lateral and ventral protrusion of a coxa, forming a novel locking mechanism and restraining a hind leg so that it is poised to extend (see supplementary information). A hind leg does not move in this cocked position, so the energy generated by an almost isometric contraction of the trochanteral depressor muscle could be stored in the strengthened thoracic cuticle, which may contain resilin^{8,9}, or in its massive and elaborate tendon.

For a hind leg to extend, the femur must disengage with the coxa, probably because the force generated by the trochanteral depressor muscle exceeds the restraining ability of the lock. The femoral ridge then suddenly snaps past the coxal protrusion, allowing the stored force to power a rapid depression of the coxo-trochanteral joint at angular velocities of $75,500$ deg s⁻¹ and a concomitant full extension of the femoro-tibial joint. The time taken from the first visible leg movements to the insect becoming airborne is no more than 1 ms.

For their size, froghoppers outperform other insects: they exceed the height jumped by the flea relative to body length and accelerate their much heavier bodies four times faster. The force exerted is 414 times their body weight and is therefore much higher than in other jumpers such as fleas (135 times)³, locusts (8 times)⁴ and humans (2–3 times)¹⁰. Froghoppers have relatively short and light hind legs that are powered by huge jumping muscles and a novel locking mechanism that allows force generated before the jump to be released rapidly. During horizontal walking, this specialization is such that the hind legs are merely dragged along while they are poised for jumping.

Malcolm Burrows

Department of Zoology, Downing Street, University of Cambridge, Cambridge CB2 3EJ, UK
e-mail: mb135@hermes.cam.ac.uk

- Alexander, R. M. *Phil. Trans. R. Soc. Lond. B* **347**, 235–248 (1995).
- Burrows, M. & Morris, O. J. *Exp. Biol.* **206**, 1035–1049 (2003).
- Bennet-Clark, H. C. & Lucey, E. C. A. *J. Exp. Biol.* **47**, 59–76 (1967).
- Bennet-Clark, H. C. *J. Exp. Biol.* **63**, 53–83 (1975).
- Evans, M. E. G. *J. Zool. Lond.* **169**, 181–194 (1973).
- Bennet-Clark, H. C. in *Perspectives in Experimental Biology* (ed. Spencer Davies, P.) 467–479 (Pergamon, Oxford, 1976).
- Weis-Fogh, T. *J. Exp. Biol.* **33**, 668–684 (1956).
- Sander, K. *Zool. Jb. Jena (Anat.)* **75**, 383–388 (1957).
- Rothschild, M., Schleim, J., Parker, K., Neville, C. & Sternberg, S. *Phil. Trans. R. Soc. Lond. B* **271**, 499–515 (1975).
- Dowling, J. J. & Vámos, L. *J. Appl. Biomechan.* **9**, 95–110 (1993).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.

Fibre optics

Surgery by sunlight on live animals

Even though the deployment of solar radiation for surgery^{1–3} must be restricted to clear-sky periods in sun-belt climates, its appeal lies in its potentially low cost compared with conventional laser fibre-optic treatments⁴. Here we show that interstitial fibre-optic solar surgery can be used effectively to kill tissue in live animals, with highly concentrated sunlight producing the same rapid, localized and extensive damage that is achieved in laser surgery. To our knowledge, this is the first time that intense incoherent light has been applied successfully in an interstitial medical procedure to kill a sizeable and prescribed extent of organ tissue photothermally.

The key factor in most photothermal fibre-optic surgery is not the coherence or monochromaticity of the light, but attaining sufficiently high power density. Solar radiation is particularly effective because its average optical-penetration depth is of the order of millimetres; also, it does not preclude the realization of ultra-high power densities in concentrators⁵.

Our prototype solar fibre-optic concentrator decouples the collection and delivery of high-density photon flux. A compact outdoor optical system transports several watts of solar radiation into the operating theatre through a flexible, high-transmissivity optical fibre of numerical aperture 0.66 and up to 20 m in length, at a power density commensurate with that of lasers used for surgery (of the order of several watts per mm²)^{1,3,5}. *Ex vivo* experiments on chicken breasts and livers indicate that a similar type, rate and extent of tissue transformation can be achieved by solar and laser surgery^{2,3}.

This comparable performance of solar and laser fibre-optic surgery stems from their equivalent optical and biophysical properties. The advantages of solar surgery over radiofrequency ablation, cryoablation and electrocautery are therefore the same as those for laser surgery in comparison with these techniques⁶. However, an advantage of solar compared with laser surgery, apart from cost, is that it does not carry the risk of eye injury to the operating team because concentrated light is delivered over a large angular range¹.

For these pilot clinical trials on live animals, our aim was to necrose (by means of coagulation and ablation) a roughly hemispherical liver lesion of about 1 cm³ volume (which corresponds to the size of an average tumour in rats), within a period of 100 s or so, by delivering 2–3 watts of highly concentrated solar radiation.

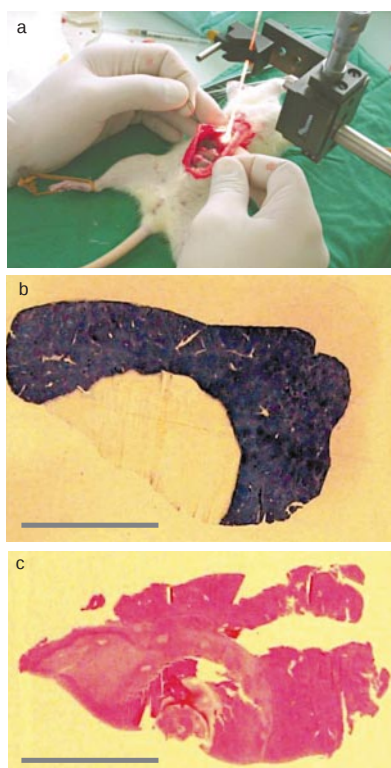


Figure 1 Solar surgery on a living rat, and the pathology of its liver after coagulation and ablation of a 12-mm section of tissue.

a, Highly concentrated sunlight being injected into the liver of a live anaesthetized rat by using an optical fibre threaded from outdoors into the operating theatre. **b**, Transverse section of the liver lesion after staining for active NADH-dependent enzymes. Viable tissue is stained blue; the surgical lesion appears as a roughly hemispherical, transparent section. **c**, Transverse section of the same sample stained with haematoxylin and eosin, which also reveals dead tissue. The hepatocyte architecture is seen to be disrupted in the necrosed region. Experiments were carried out in accordance with the Helsinki Committee of the Soroka Medical Center and Ben-Gurion University, Beersheva, Israel. Scale bars, 10 mm.

A general anaesthetic was administered to two healthy 200-g female rats. The liver was exposed to view by scalpel incision on the animal's underside. Solar radiation — whose power was measured as 2.0–2.5 watts — was injected into the upper region of the liver from the distal tip of an optical fibre of 1.0 mm diameter (Fig. 1a). Each animal was irradiated twice, in two separate sections of its liver, for periods ranging from 40 to 180 s. The diameter of the resulting lesion was measured as about 12 mm immediately after the irradiation procedure.

The rats were revived and continued to function without complication. They were

killed 24 and 72 h later, respectively. Their liver pathology was investigated immediately with a specific stain for active NADH-dependent enzymes, which therefore identifies living tissue (Fig. 1b), and with haematoxylin-and-eosin stain, which stains both living and dead cells (Fig. 1c). The NADH-enzymatic stain therefore provides superior delineation of the necrosed region of each liver.

Longitudinal and transverse cuts revealed that necrosis was roughly hemispherical and symmetric. The lesion size after 24 h (about 1 cm³) was almost twice that recorded photographically at surgery, but did not increase after 72 h. These findings are consistent with lesion sizes recorded after liver surgery by laser, which increase progressively and markedly for about 24 h after treatment⁷.

Our results indicate that the efficacy of interstitial fibre-optic solar surgery — that is, the lesion volume per unit of injected light energy — is up to 5 mm³ per joule, which is comparable to that of laser surgery when used in similar procedures.

Jeffrey M. Gordon*†, **Daniel Feuermann***, **Mahmoud Huleihil***, **Solly Mizrahi‡**, **Ruthy Shaco-Levy§**

**Department of Solar Energy and Environmental Physics, Jacob Blaustein Institute for Desert Research, Ben-Gurion University of the Negev, Sede Boqer 84990, Israel*

†Pearlstone Center for Aeronautical Engineering Studies, Department of Mechanical Engineering, and Departments of ‡Surgery A and §Pathology, Soroka Medical Center, and Faculty of Health Sciences, Ben-Gurion University of the Negev, Beersheva 84105, Israel

e-mail: jeff@bgumail.bgu.ac.il

1. Feuermann, D. & Gordon, J. M. *Opt. Eng.* **37**, 2760–2767 (1998).
2. Gordon, J. M., Feuermann, D. & Huleihil, M. *Appl. Phys. Lett.* **81**, 2653–2655 (2002).
3. Gordon, J. M., Feuermann, D., Huleihil, M., Mizrahi, S. & Shaco-Levy, R. *J. Appl. Phys.* **93**, 4843–4851 (2003).
4. Katzir, A. *Lasers and Optical Fibers in Medicine* (Academic, San Diego, 1993).
5. Feuermann, D., Gordon, J. M. & Huleihil, M. *Solar Energy* **72**, 459–472 (2002).
6. De Sanctis, J. T., Goldberg, S. N. & Mueller, P. R. *Cardiovasc. Intervent. Radiol.* **21**, 273–296 (1998).
7. Fujitomi, Y. *et al. Lasers Surg. Med.* **24**, 14–23 (1999).

Competing financial interests: declared none.

erratum

Vortex rings in a constant electric field

David G. Grier

Nature **424**, 267–268 (2003)

Yilong Han, the first author, was omitted from the list of authors of this communication. The correct listing should therefore be: Yilong Han, David G. Grier.

brief communications is intended to provide a forum for brief, topical reports of general scientific interest and for technical discussion of recently published material of particular interest to non-specialist readers (**communications arising**). Priority will be given to contributions that have fewer than 500 words, 10 references and only one figure. Detailed guidelines are available on *Nature's* website (www.nature.com/nature).

Transformation and control of ultra-short pulses in dispersion-engineered photonic crystal fibres

W. H. Reeves¹, D. V. Skryabin¹, F. Biancalana¹, J. C. Knight¹, P. St. J. Russell¹, F. G. Omenetto², A. Efimov³ & A. J. Taylor³

¹Optoelectronics Group, Department of Physics, University of Bath, Bath BA2 7AY, UK

²Physics Division, P-23 and ³Materials Science and Technology Division, MST-10, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA

Photonic crystal fibres (PCFs) offer greatly enhanced design freedom compared to standard optical fibres. For example, they allow precise control of the chromatic dispersion (CD) profile—the frequency dependence of propagation speed—over a broad wavelength range. This permits studies of nonlinear pulse propagation in previously inaccessible parameter regimes. Here we report on spectral broadening of 100-fs pulses in PCFs with anomalously flat CD profiles. Maps of the spectral and spatio-temporal behaviour as a function of power show that dramatic conversion (to both longer and shorter wavelengths) can occur in remarkably short lengths of fibre, depending on the magnitude and shape of the CD profile. Because the PCFs used are single-mode at all wavelengths, the light always emerges in a fundamental guided mode. Excellent agreement is obtained between the experimental results and numerical solutions of the nonlinear wave equation, indicating that the underlying processes can be reliably modelled. These results show how, through appropriate choice of CD, nonlinearities can be efficiently harnessed to generate laser light at new wavelengths.

A feature common to almost all pulses of energy is that they broaden (or ‘disperse’) as they travel. Thus it was a surprise when, in 1844, Russell observed a solitary wave of water travelling, without change in shape, along a canal near Edinburgh¹. Solitary waves turn out to be quite common in nature. Another example is of tsunamis—giant water waves initiated by seismic shocks and then transmitted across oceans in the form of massive pulses of energy. How does this happen? Fourier analysis shows that pulses can be decomposed into a spectrum of sinusoidal waves spread over a finite frequency band. Chromatic dispersion (CD)—a linear effect that affects most types of wave—causes different frequencies to travel at different speeds, producing pulse broadening. Solitary waves form when the effects of CD are balanced by nonlinearity—a phenomenon that changes the properties of systems at high power densities.

In 1971, Zakharov and Shabat² showed theoretically that, for an appropriate combination of nonlinearity and CD, light can form solitary waves, or solitons, which are robust against any perturbations, including strong inter-soliton collisions. The light solitons exist because self-phase modulation (a nonlinear effect caused by the tendency for the wavefront velocity to fall as the intensity increases) gradually redistributes the pulse energy into frequencies above and below the pulse’s central frequency. A positive frequency ‘chirp’ develops across the broadening pulse: lower frequencies arrive first. On the other hand, CD broadening also gives rise to a frequency chirp because the different spectral components travel at different ‘group’ velocities v_G :

$$D(\lambda) = \frac{\partial}{\partial \lambda} \frac{1}{v_G} = -\frac{2\pi c}{\lambda^2} \frac{\partial}{\partial \omega} \frac{1}{v_G} = -\frac{2\pi c}{\lambda^2} \beta_2(\omega) \neq 0 \quad (1)$$

where $D(\lambda)$ and $\beta_2(\omega)$ represent the CD as a function of wavelength and frequency, λ is the wavelength in vacuum, ω the angular frequency and c the speed of light in vacuum. In the anomalous CD region ($D(\lambda) > 0$), ‘red’-shifted frequencies travel more slowly than ‘blue’-shifted ones, producing a negative frequency chirp that can exactly balance the positive chirp of self-phase modulation. When this happens, a soliton forms. In 1973 Hasegawa suggested that solitons could appear in optical fibres in the wavelength range beyond $\sim 1,300$ nm where the CD is anomalous³, and then in 1980

Mollenauer went on to report the first experimental observation of optical solitons in fibres⁴. Since then, many optical communications systems based on solitons have been developed, including most recently so-called ‘dispersion-managed’ systems where the CD changes sign periodically along the length⁵ and systems where femtosecond pulses are adaptively controlled⁶.

This story of prediction, exploration and discovery has been told and retold many times in textbooks⁷. But to what degree can the behaviour of solitons, and other related nonlinear effects, be engineered? Fibre nonlinearity is determined by the nonlinear susceptibility of the silica glass and the effective cross-sectional area of the guided mode A_{eff} . It is represented by the parameter γ (in $\text{m}^{-1} \text{W}^{-1}$):

$$\gamma = \frac{\omega}{c} \frac{n_2}{A_{\text{eff}}} = \frac{2\pi}{\lambda} \frac{n_2}{A_{\text{eff}}} \quad (2)$$

where n_2 is the nonlinear refractive index ($\approx 2.2 \times 10^{-20} \text{ m}^2 \text{W}^{-1}$ in silica glass). Although A_{eff} can be engineered to a degree, it turns out that the CD is itself strongly dependent upon the shape and size of the core. So should one concentrate on increasing the nonlinearity or controlling the CD? Recent work^{8–11} on photonic crystal fibres (PCFs) with ultra-small A_{eff} shows that control of CD profile $D(\lambda)$ is immensely more important. Previous numerical modelling and experimental work hints at this, though in fibres without precisely controlled $D(\lambda)$.

Here we report our success in refining $D(\lambda)$ to the point where new regimes of nonlinear behaviour become accessible.

Fibre design and fabrication

Underpinning the study is PCF^{12,13}. PCF consists of a strand of glass with an array of microscopic air-channels running along its length. It can guide light by two mechanisms: a photonic bandgap (when the refractive index of the core can have any value), and a modified form of total internal reflection (when the holey cladding has a lower index, on average, than the core). Over the past decade PCF has attracted increasing interest in many fields, including nonlinear fibre optics, where the ability to design PCFs with well-controlled CD profiles and high nonlinearity is leading to a number of new applications^{14–16}.

Silica glass PCF is made in several stages. First, a set of precision-manufactured glass capillaries and rods are stacked into a macroscopic 'preform' of the desired microstructure. In the fibres studied here, a hexagonal lattice of capillaries is arranged around a central solid core. This preform stack is then fused together and reduced in size (by several orders of magnitude) in a fibre-drawing tower. Even this simple structure allows remarkable control over the CD of the mode guided in the central solid core. By choosing appropriate values of hole diameter d and interhole spacing Λ , the zero CD wavelength (which is close to 1,300 nm in standard fibre) can be shifted down into the 800-nm wavelength band, allowing highly efficient supercontinuum generation using femtosecond pulses from Ti:sapphire lasers¹⁷. The shape of $D(\lambda)$ can also be flattened in the 1,550-nm (100-fs) wavelength range¹⁸, the resulting fibres having the additional advantage of being single-mode at all wavelengths where the glass is transparent (standard flattened-dispersion step-index fibres are single-mode over a much smaller range)^{19,20}. Broad-band single-mode operation has additional advantages in applications where well-controlled interactions between signals at widely different wavelengths are required, such as optical parametric amplifiers²¹ and in supercontinuum-based infrared communications²².

Experimental results

An important challenge in the design of PCFs—particularly those with flattened CD—is the absence of accurate analytical formulae for $D(\lambda)$. This means that numerical modelling must be used to provide the initial 'ideal' design (hole diameter and spacing) for a desired $D(\lambda)$ (ref. 18). Fabricating precisely this ideal structure is also a challenge, because of slight systematic distortions introduced in the fibre-drawing process. Starting from a suitable preform stack, a large number of PCFs were drawn under slightly different conditions, each yielding a different profile of flattened $D(\lambda)$, ranging from $17 \text{ ps nm}^{-1} \text{ km}^{-1}$ to $-7 \text{ ps nm}^{-1} \text{ km}^{-1}$. Figure 1 shows $D(\lambda)$ for the fibres used in the experiments, measured using low coherence interferometry²³. The plot includes $D(\lambda)$ for a PCF with anomalous CD and negative CD slope $\partial D/\partial\lambda < 0$ (fibre labelled A in Fig. 1), a PCF with near-zero anomalous CD ($\sim +2 \text{ ps nm}^{-1} \text{ km}^{-1}$) and $\partial D/\partial\lambda > 0$ (fibre B), and a PCF with normal CD and $\partial D/\partial\lambda < 0$ (fibre C). Included in the plot is a

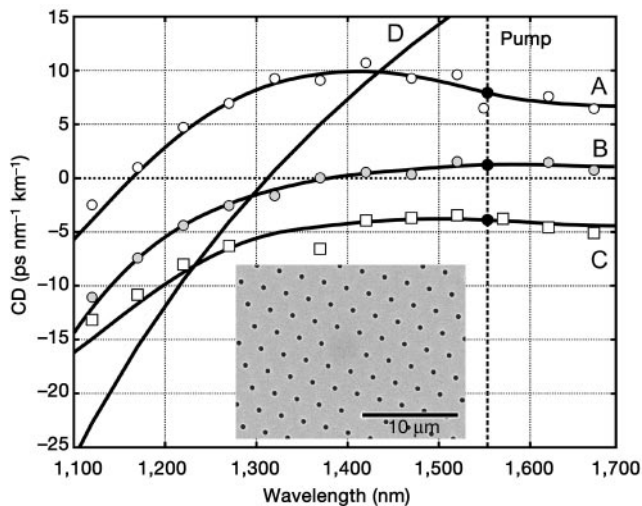


Figure 1 Dispersion profiles for three PCFs (A–C) and a Corning SMF28 fibre (D). The circles and squares are experimental data, and the curves are polynomial fits. The broken vertical line indicates the pump wavelength in the experiments. The scanning electron micrograph is of fibre B, with holes of diameter $\sim 0.6 \mu\text{m}$ and an interhole spacing of $2.6 \mu\text{m}$.

scanning electron micrograph (SEM) of fibre B. The fibre parameters, including the coefficients of a polynomial fit to the dispersion $\beta_2(\omega)$ (see equation (3) below), are given in Table 1. The small effective core-cladding refractive index difference and high structural regularity of the fibres meant that any 'accidental' birefringence in the guided mode was negligible. All of the PCFs used in the experiments had $A_{\text{eff}} \approx 44 \mu\text{m}^2$, a numerical aperture of approximately 0.2 and a measured loss of $\sim 35 \text{ dB km}^{-1}$ at 1,550 nm.

Light at 1,550 nm from an optical parametric oscillator, pumped by a Ti:sapphire laser, was launched into 1-m lengths of fibre. The pulse length was approximately 100 fs, the repetition rate was 80 MHz and a maximum pulse energy of several nanojoules was available at the fibre input face. The input power was varied using crossed polarizers and the generated spectrum was measured at the fibre output using an optical spectrum analyser (OSA). The generated spectra, plotted against launched power for the three PCFs in Fig. 1, are presented in Fig. 2. The left-hand column shows the experimental results and the right-hand column contains the predictions of the numerical model (see below).

Note that the long-wavelength portions of the generated spectra lie beyond the maximum operating wavelength of the OSA (1,650 nm), and that wavelength-dependent losses, caused by imperfect coupling of light into the OSA, may also distort the recorded spectra to some degree.

Modulational instability

When a strictly monochromatic wave of high power is travelling in a dispersive nonlinear medium, instabilities can appear within certain frequency bands on either side of the pump frequency. This causes signals to grow from noise, and is at the root of most nonlinear effects. We start with a system whose CD is given by:

$$\beta_2(\omega) = \sum_{m=2}^{\infty} \frac{\beta_m(\omega_0)}{(m-2)!} \Omega^{m-2} \quad (3)$$

where $\Omega = (\omega - \omega_0)$ and ω_0 is the pump frequency (if the coefficients in Table 1 are used, Ω is measured in units of $2\pi \text{ THz}$). By extending the perturbation analysis in ref. 7 and taking terms up to $m = 6$, it may be shown that the modulational instability (MI) gain per metre has the value:

$$\gamma_{\text{MI}} = \text{Im} \left[\sqrt{Q(Q + 2/L_{\text{NL}})} \right], \quad (4)$$

$$Q = \beta_2 \Omega^2 / 2 + \beta_4 \Omega^4 / 24 + \beta_6 \Omega^6 / 720$$

where $L_{\text{NL}} = (\gamma P)^{-1}$ is the nonlinear length and P is the pump power. The frequencies at which γ_{MI} becomes real-valued are located at the zeroes of the function under the square-root in equation (4). Note that odd CD orders play no role in the presence or absence of gain. This is because they cancel out from the degenerate four-wave phase-matching condition $2\beta(\omega_0) - \beta(+\Omega) - \beta(-\Omega) = 0$ where $\beta(\omega)$ is the exact wavevector (including dispersion). The results are summarized in density plots of gain versus Ω and β_2 for four different cases (Fig. 3). Standard textbook analyses, treating the case when $\beta_{n>2} = 0$, show that gain occurs only when $\beta_2 \leq 0$. The situation is much more complicated when higher-order CD terms are present. Gain can appear in the normal dispersion region $\beta_2 > 0$, sometimes in several sidebands. In the

Table 1 Hole size and spacing for the three PCFs used

Fibre	d/Λ	$\Lambda (\mu\text{m})$	β_2	β_3	$\beta_4 \times 10^4$	$\beta_5 \times 10^9$	$\beta_6 \times 10^{10}$
A	0.26	2.35	-11.2	-0.0044	+1.06	-6.3	-1.09
B	0.22	2.41	-1.55	+0.0065	+0.61	-104	+1.01
C	0.24	2.00	+4.97	-0.015	+0.81	-62.9	-0.34

The values of the coefficients in the polynomial fits to the CD profiles (units are $\text{ps}^2 \text{ km}^{-1}$ where n is the order of β_n) are also given. The hole sizes and spacings were measured close to the core.

presence of $\beta_4 < 0$ it is interesting that the gain bands broaden out. These results illustrate the unusual nonlinear effects that can be seen in fibres with non-standard CD profiles, such as those in which $\beta_2 \equiv 0$.

Numerical modelling

The equation used to model the experiment is the so-called generalized nonlinear Schrödinger equation:

$$i\partial_z A + \sum_{m \geq 2} \frac{i^m L_{D2}}{m! L_{Dm}} \text{sgn}(\beta_m) \partial_t^m A + N^2 \left(1 + \frac{i}{\omega_0 t_0} \partial_t \right) A \int_{-\infty}^{+\infty} R(q) |A(t-q, z)|^2 dq = 0 \quad (5)$$

In this equation $t_0 = 0.1$ ps is the full-width at half-maximum (FWHM) pulse duration, ω_0 is the pump frequency, $\beta_m = \partial^m(1/\nu_G)/\partial \omega^m$ (evaluated at $\omega = \omega_0$) is the m th dispersion coefficient,

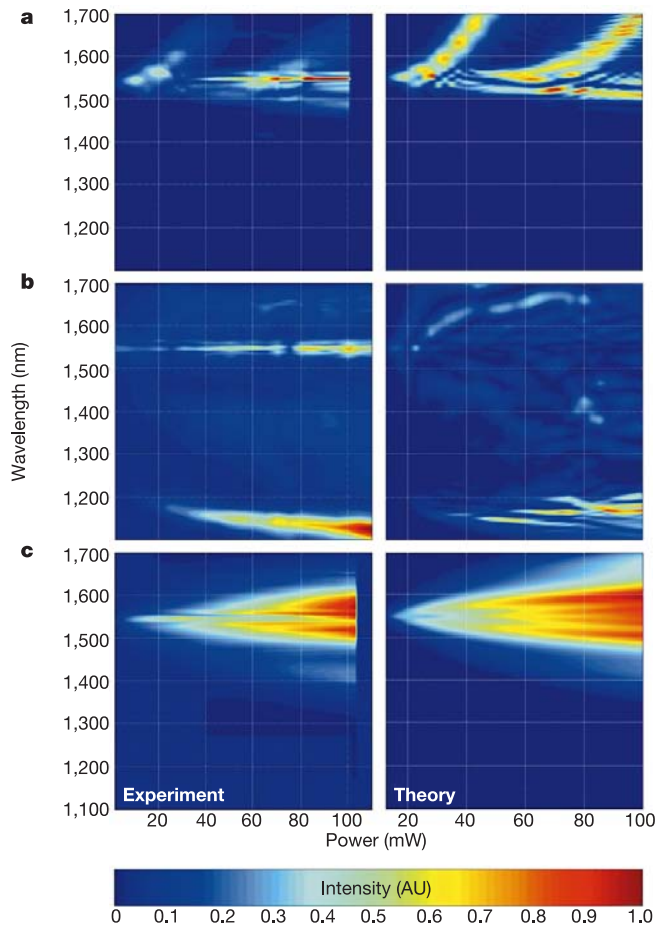


Figure 2 Experimental (left) and theoretical (right) output intensity spectra for the PCFs in Fig. 1, as a function of the average power. The pump source is a pulse train of 100-fs pulses at a repetition rate of 80 MHz and a wavelength of 1,550 nm in all cases. Plots show **a**, fibre A, with anomalous CD and negative slope; **b**, fibre B, with small anomalous CD and positive slope; and **c**, fibre C, with normal CD and negative slope. The theoretical modelling was based on the experimental data, but without one of the fitting parameters—to allow slight rescaling of the power axis (see text). The small disparities between experiment and theory are caused by limitations in the optical measurements, inexact knowledge of the CD profile at all wavelengths, and the perfect chirp-free hyperbolic-secant time-envelopes used to represent the input pulses.

$L_{Dm} = t_0^m / |\beta_m|$ is the m th dispersion length, and

$$R(t) = (1 - f_R) \delta(t) + f_R \frac{\tau_1^2 + \tau_2^2}{\tau_1 \tau_2} e^{-t/\tau_2} \sin\left(\frac{t}{\tau_1}\right) \Theta(t) \quad (6)$$

is the response function, which includes instantaneous electronic and delayed Raman contributions, the causality condition being taken into account through the Heaviside function $\Theta(t)$ (unity for $t > 0$ and zero for $t < 0$). The time t is in a reference frame moving at the group velocity (evaluated at $\omega = \omega_0$) and is measured in the units of t_0 . The propagation distance z along the fibre is measured in units of L_{D2} . N^2 is the ratio between the peak power of the pump pulse and the peak power $P_0 = (\gamma L_{D2})^{-1}$ necessary to create a single soliton in the ideal nonlinear Schrödinger equation (that is, equation (5), neglecting the Raman contribution and dispersion terms of order $m > 2$). For all the PCFs modelled, $\gamma \approx 0.002 \text{ m}^{-1} \text{ W}^{-1}$. A is the amplitude of the electric field at the pump frequency, measured in the units of $\sqrt{P_0}$. The parameters of the response function were taken from ref. 7: $f_R = 0.18$, $\tau_1 = 12.2$ fs/ t_0 and $\tau_2 = 32$ fs/ t_0 .

It is generally accepted that equation (5) accurately describes the propagation of femtosecond pulses and that it is valid even in circumstances where the bandwidth of the radiation is of the same order as the central frequency of the pump laser—as confirmed in recent modelling of supercontinuum generation in PCF (see, for example, ref. 24).

To solve equation (5) we used a fast Fourier transform method to integrate the linear part, and a second-order Runge–Kutta algorithm for the nonlinear part. The launched pulses were taken to have hyperbolic-secant temporal profiles and to be free of any frequency chirp, which closely approximates the experimental condition where the launched pulses exhibit a minimal linear chirp. To convert the average pump power P_{av} (measured in experiments) to peak power, we used the relation $P_{av} = \nu_{fsr} \eta t_0 (P_0 N^2)$, where $\nu_{fsr} = 80$ MHz is the pulse repetition rate and $\eta = 0.7$ is an empirically determined coefficient which is introduced to allow for non-ideal input pulse profiles (this affects the relationship between peak and average power).

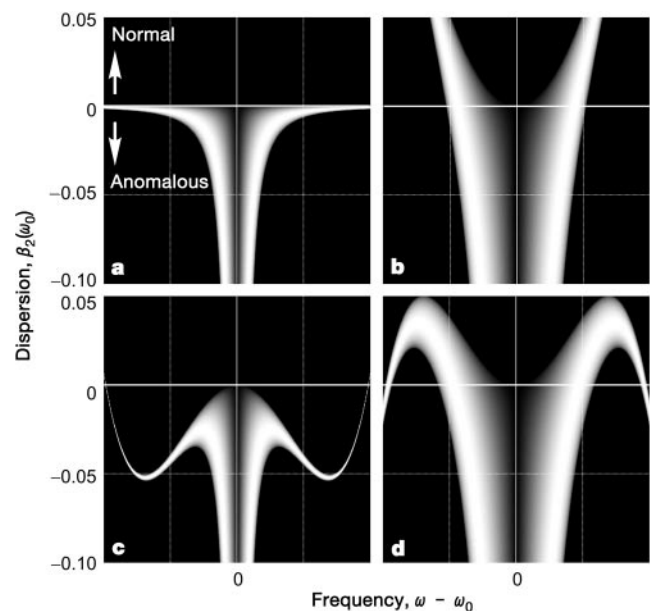


Figure 3 Modulational instability gain in systems with different $D(\lambda)$ (the grey to white regions represent medium to high gain on a linear scale; gain is zero in the black regions). The parameter values in each case are $L_{NL} = 0.1$ with: **a**, $\beta_4 = 0$ and $\beta_6 = 0$; **b**, $\beta_4 = -10^{-3}$ and $\beta_6 = 0$; **c**, $\beta_4 = 10^{-3}$ and $\beta_6 = -1.24 \times 10^{-5}$; and **d**, $\beta_4 = -10^{-3}$ and $\beta_6 = 1.24 \times 10^{-5}$. The units in each case are $\text{ps}^n \text{ km}^{-1}$ for β_n .

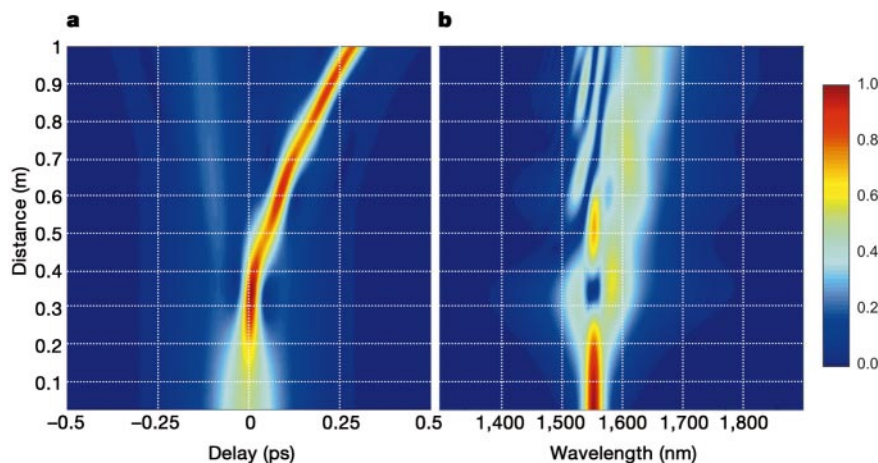


Figure 4 The pulse evolution, plotted against propagation distance, for fibre A in Fig. 1: **a**, Pulse delay (relative to a time frame moving at the pulse's average group velocity); **b**, spectrum. The colour represents a uniform mapping to an arbitrary intensity scale.

Neglecting fibre losses and launching chirp-free pulses will obviously create some discrepancies between the numerical and experimental results. More substantial difficulties arise from the fact that $D(\lambda)$ was measured only in the range 1 to 1.7 μm , whereas the numerical modelling produces spectra that can cover a much broader range. To calculate the dispersion parameters for equation (5), a 4th-order polynomial ($m = 6$) was fitted to the available experimental CD profile. For each fibre type, the coefficients in the summation in equation (3) are given in Table 1. Adding terms beyond the 4th order did not lead to any noticeable change in the modelled behaviour.

Results and discussion

Soliton formation is expected in PCFs with appreciable flattened anomalous dispersion. This indeed is quite clearly observed in the numerical modelling of fibre A. Figure 4a shows evolution of the pulse intensity in the (t, z) plane for a pump power of $P_{\text{av}} = 48$ mW. Initially self-phase modulation causes pulse compression and later, when dispersive effects come into play, one can clearly see the formation and splitting away of the solitary pulse, which is delayed (that is, continuously shifted towards larger values of t) by the Raman effect. Figure 4b shows spectral evolution of the same initial pulse, showing that the solitary pulse is red-shifted, as is typical for Raman solitons. The pulse dynamics in this particular PCF are similar to those observed in standard telecommunication fibres.

The only noticeable difference is that nonlinear effects, including the Raman part of the nonlinearity, are stronger in the PCF case, which leads to the formation of a Raman-shifted solitary wave at an earlier stage.

Our experimental measurements in this and all other cases show the dependence of output spectrum on the average input power. Comparing experimental and numerical results in Fig. 2a, good qualitative agreement is seen. In particular, two clear spectral traces deviating out towards longer wavelengths are present in both experiment and theory. These traces correspond to the formation of Raman-delayed solitary waves; most of the non-solitonic radiation continues to propagate at the initial pump wavelength. Although the modelling results are in qualitative agreement with experiment, there are differences in the relative brightness of the observed and predicted spectral lines, especially in the red part of the spectrum. This is most probably due to the combination of the uneven spectral response of the OSA and to the wavelength-dependent coupling losses into the instrument.

Modelling of pulse propagation in fibre B—the one with small anomalous CD—is shown in Fig. 2b and Fig. 5. The behaviour in this case is highly sensitive to the exact form of $D(\lambda)$. Nevertheless there is good qualitative agreement between experiment and theory. After the initial stages of propagation, which is again dominated by the spectral broadening and temporal compression induced by self-phase modulation, a quasi-solitary pulse forms (Fig. 5a). This pulse

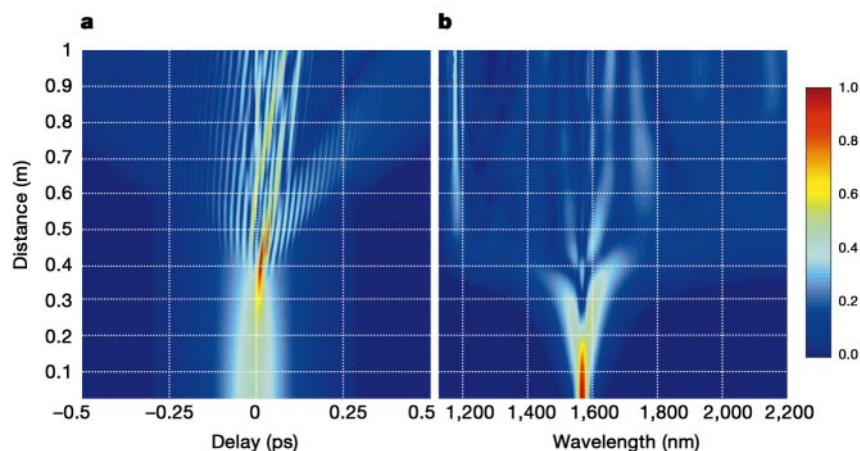


Figure 5 The pulse evolution, plotted against propagation distance, for fibre B in Fig. 1: **a**, Pulse delay (relative to a time frame moving at the pulse's average group velocity); **b**, spectrum. The colour represents a uniform mapping to an arbitrary intensity scale.

immediately starts to radiate dispersive waves (waves that are too weak to be affected by nonlinearity). Unlike in the previous case (Fig. 4a), this radiation is strong enough to prevent formation of a robust solitary pulse. Another noticeable difference, compared to the previous case, is that spectral broadening now happens on both sides of the pump wavelength. In particular, one can clearly see formation of a robust bright feature at short wavelengths. This blue-shifted radiation, observed in both modelling and experiment, is due to the effect of so-called resonant radiation²⁵.

Resonant radiation can be understood as follows. The quasi-soliton formed as a result of self-phase modulation can be approximately considered as almost non-dispersive pulse. Dispersion characteristic of such a pulse is a straight line in the wavevector-frequency plane with a slope given by the pulse group velocity. In contrast, small-amplitude linear waves are highly dispersive, their wavevectors being complicated functions of frequency. Resonant radiation occurs when one or more spectral components of the soliton are phase-matched to the dispersive waves, that is, their wavevectors are equal. Thus intersections of the two dispersion characteristics guarantee existence of the channels for resonant energy exchange between solitons and the 'sea' of ever-present linear dispersive waves. It results in the coherent amplification of phase-matched linear waves. In the case shown in Fig. 2b there are two resonant frequencies: one is on the blue side of the soliton spectrum and the other one is on the red. However, the red-shifted line is far detuned and therefore very weak, and hardly detectable in either experiment or modelling.

Finally, in Fig. 2c, the CD is normal everywhere and spectral broadening is caused by the higher-order dispersion terms. In this case, solitary wave formation, which takes place in both of the previous cases with $\beta_2 < 0$, is absent. The excellent agreement between modelling and experiment arises from the fact that practically no energy is transferred into spectral areas where the CD is not known.

Future prospects

The ability to control light on the femtosecond timescale, by manipulating the CD profile precisely over a broad wavelength band, defines a new territory in nonlinear optics. The implications for laser physics and its applications are wide-reaching. For example, it is difficult (and expensive) to generate laser light at wavelengths beyond those offered by the small canon of available lasers. The efficient transfer of laser energy to new wavelengths is therefore a highly attractive prospect.

In Fig. 2a energy is transferred to longer wavelengths in a highly controllable way; in Fig. 2b there is conversion to a dramatically shorter wavelength—and the magnitude of the wavelength shift can be engineered. In Fig. 2c the spectrum broadens in a symmetrical manner, and at higher pump powers a supercontinuum is generated that can extend from the ultraviolet (~ 350 nm) to the near-infrared ($\sim 2,000$ nm) (ref. 14), with a spectral brightness 10,000 times greater than sunlight. By adjusting the CD profile, supercontinua can be efficiently generated from the inexpensive, ultracompact, all-solid-state pulsed lasers that are becoming available, for example, at 1,064 and 1,550 nm, with pulse durations from ~ 10 fs to ~ 600 ps. Further control can be achieved by pre-chirping and pre-shaping the pump pulses²⁶, or by exploiting nonlinear coupling between orthogonal polarization states of the light in birefringent PCFs (ref. 8).

An additional degree of freedom comes from the use of borosilicate or non-silica glasses²⁷, where the intrinsic CD profile of the material is quite different and transparency is possible in new wavelength ranges, for example, out to $10\text{ }\mu\text{m}$ in the infrared, where there is a huge unsatisfied demand for tunable laser sources, driven largely by the requirement to sense hydrocarbons (such as methane and atmospheric pollutants) with strong molecular absorptions in the range 2 to $10\text{ }\mu\text{m}$. Fibre delivery of high-energy femtosecond laser pulses for biomedical sensing and laser manu-

facturing (cutting and drilling of materials) requires precise control of CD—otherwise the pulses broaden before they arrive.

Better control of the CD profile will allow generation of spectrally flat supercontinua, which is highly desirable in many applications from optical coherence tomography¹⁶ to frequency metrology¹⁵. The unique combination of extremely high spatial resolution ($\sim 1\text{ }\mu\text{m}$) with huge spectral breadth and high brightness makes the characterization of diffractive structures—such as those that cause structural colour in nature²⁸—quick and accurate. Finally, control of the dispersion profile in hollow-core gas-filled PCF will allow all sorts of super-efficient nonlinear devices for wavelength conversion, such as gas-Raman cells²⁹. CD-engineered PCF seems set to revolutionize many areas of nonlinear laser science and technology and their applications. □

Received 9 April; accepted 4 June 2003; doi:10.1038/nature01798.

- Russell, J. S. Report on waves. in *Report of the 14th Meeting of the British Association for the Advancement of Science* 311–390, Plates XLII–LVII (London, 1845).
- Zakharov, V. E. & Shabat, A. B. Exact theory of two-dimensional self-focusing and one-dimensional self-modulation of nonlinear waves in nonlinear media. *Sov. Phys. JETP* **34**, 62–69 (1971).
- Hasegawa, A. & Tappert, F. D. Transmission of stationary nonlinear optical pulses in dispersive dielectric fibres. *Appl. Phys. Lett.* **23**, 142–144 (1973).
- Mollenauer, L. E., Stolen, R. H. & Gordon, J. P. Experimental observation of picosecond pulse narrowing and solitons in optical fibers. *Phys. Rev. Lett.* **45**, 1095–1098 (1980).
- Smith, N. J., Knox, F. M., Doran, N. J., Blow, K. J. & Bennion, I. Enhanced power solitons in optical fibres with periodic dispersion management. *Electron. Lett.* **32**, 54–55 (1996).
- Omenetto, F. G., Moores, M. D., Reitze, D. H. & Taylor, A. J. Adaptive control of nonlinear femtosecond pulse propagation in optical fibers. *Opt. Lett.* **26**, 938–940 (2001).
- Agrawal, G. P. *Nonlinear Fiber Optics* (Academic, San Diego, 2001).
- Ortígoza-Blanch, A., Knight, J. C. & Russell, P. St. J. Pulse breaking and supercontinuum generation with 200-fs pump pulses in photonic crystal fibers. *J. Opt. Soc. Am. A* **19**, 2567–2572 (2002).
- Herrmann, J. et al. Experimental evidence for supercontinuum generation by fission of higher-order solitons in photonic fibers. *Phys. Rev. Lett.* **88**, 173901 (2002).
- Coen, S. et al. Supercontinuum generation by stimulated Raman scattering and parametric four-wave mixing in photonic crystal fibers. *J. Opt. Soc. Am. B* **19**, 753–764 (2002).
- Gaeta, A. L. Nonlinear propagation and continuum generation in microstructured optical fibers. *Opt. Lett.* **27**, 924–926 (2002).
- Knight, J. C., Birks, T. A., Russell, P. St. J. & Atkin, D. M. All-silica single-mode fiber with photonic crystal cladding. *Opt. Lett.* **21**, 1547–1549 (1996); erratum *Opt. Lett.* **22**, 484–485 (1997).
- Russell, P. St. J. Photonic crystal fibers. *Science* **299**, 358–362 (2003).
- Jones, D. A. et al. Carrier-envelope phase control of femtosecond mode-locked lasers and direct optical frequency synthesis. *Science* **288**, 635–639 (2000).
- Holzwarth, R. et al. An optical frequency synthesiser for precision spectroscopy. *Phys. Rev. Lett.* **85**, 2264–2267 (2000).
- Hartl, X. D. et al. Ultrahigh resolution optical coherence tomography using continuum generation in an air-silica microstructure optical fiber. *Opt. Lett.* **26**, 608–610 (2001).
- Wadsworth, W. J. et al. Supercontinuum generation in photonic crystal fibers and optical fiber tapers: a novel light source. *J. Opt. Soc. Am. B* **9**, 2148–2155 (2002).
- Reeves, W. H., Knight, J. C., Russell, P. St. J. & Roberts, P. J. Demonstration of ultra-flattened dispersion in photonic crystal fibers. *Opt. Expr.* **10**, 609–613 (2002).
- Lundin, R. Dispersion flattening in a W-fibre. *Appl. Opt.* **6**, 1011–1014 (1994).
- Lee, J., Song, G. H., Paek, U.-C. & Seo, Y. G. Design and fabrication of a nonzero-dispersion fibre with a maximally flat dispersion spectrum. *IEEE Phot. Tech. Lett.* **4**, 317–319 (2001).
- Hansryd, J. & Andrekson, P. A. Broad-band continuous-wave-pumped fiber optical parametric amplifier with 49 dB gain and wavelength-conversion efficiency. *IEEE Phot. Tech. Lett.* **13**, 194–196 (2001).
- Takushima, Y. & Kikuchi, K. 10-GHz, over 20-channel multiwavelength pulse source by slicing supercontinuum spectrum generated in normal-dispersion fiber. *IEEE Phot. Tech. Lett.* **11**, 322–324 (1999).
- Tateda, M., Shibata, N. & Seikai, S. Interferometric method for chromatic dispersion measurement in a single-mode optical fiber. *IEEE J. Quant. Electron.* **17**, 404–407 (1981).
- Dudley, J. M. et al. Supercontinuum generation in air-silica microstructured fibers with nanosecond and femtosecond pulse pumping. *J. Opt. Soc. Am. B* **19**, 765–771 (2002).
- Akhmediev, N. & Karlsson, M. Cerenkov radiation emitted by solitons in optical fibers. *Phys. Rev. A* **51**, 2602–2607 (1995).
- Weiner, A. Femtosecond pulse shaping using spatial light modulators. *Rev. Sci. Instrum.* **71**, 1929–1960 (2000).
- Ravi Kanth Kumar, V. V. et al. Extruded soft glass photonic crystal fiber for ultrabroad supercontinuum generation. *Opt. Expr.* **10**, 1520–1525 (2002).
- Parker, A. R. 515 million years of structural colour. *J. Opt. A: Pure Appl. Opt.* **2**, R15–R28 (2000).
- Benabid, F., Knight, J. C., Antonopoulos, G. & Russell, P. St. J. Stimulated Raman scattering in hydrogen-filled hollow-core photonic crystal fiber. *Science* **298**, 399–402 (2002).

Acknowledgements We thank B. J. Mangan of BlazePhotonics Ltd for technical assistance in fabricating the fibres.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.C.K. (j.c.knight@bath.ac.uk) or F.O. (omenetto@lanl.gov).

Integration of interferon- α/β signalling to p53 responses in tumour suppression and antiviral defence

Akinori Takaoka¹, Sumio Hayakawa¹, Hideyuki Yanai¹, Dagmar Stoiber^{1*}, Hideo Negishi¹, Hideaki Kikuchi¹, Shigeru Sasaki², Kohzoh Imai², Tsukasa Shibue¹, Kenya Honda¹ & Tadatsugu Taniguchi¹

¹Department of Immunology, Faculty of Medicine and Graduate School of Medicine, University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan

²First Department of Internal Medicine, Sapporo Medical University School of Medicine, South 1, West 17, Chuo-ku, Sapporo 060-8556, Japan

* Present address: Institute of Pharmacology, University of Vienna, Waehringerstrasse 13a, A-1090 Vienna, Austria

Swift elimination of undesirable cells is an important feature in tumour suppression and immunity. The tumour suppressor p53 and interferon- α and - β (IFN- α/β) are essential for the induction of apoptosis in cancerous cells and in antiviral immune responses, respectively, but little is known about their interrelationship. Here we show that transcription of the *p53* gene is induced by IFN- α/β , accompanied by an increase in p53 protein level. IFN- α/β signalling itself does not activate p53; rather, it contributes to boosting p53 responses to stress signals. We show examples in which *p53* gene induction by IFN- α/β contributes to tumour suppression. Furthermore, we show that p53 is activated in virally infected cells to evoke an apoptotic response and that p53 is critical for antiviral defence of the host. Our study reveals a hitherto unrecognized link between p53 and IFN- α/β in tumour suppression and antiviral immunity, which may have therapeutic implications.

The tumour suppressor p53, activated in response to DNA damage, induces cell cycle arrest or apoptosis through transcriptional activation of its target genes, hence having a central role in tumour suppression^{1–5}. So far, it is not known whether p53 contributes to the immune responses that lead to the eradication of pathogens such as viruses. On the other hand, IFN- α/β , both of which are essential cytokines for antiviral immunity, are sometimes referred to as 'negative growth factors', and manifest anti-oncogenic activities^{6–10}. In fact, IFN- α/β are used for the treatment of some forms of human cancer but the molecular basis for the treatment is poorly understood^{11–13}. Until now, little if anything has been known about the link between the p53 and IFN- α/β system.

Induction of p53 protein by IFN- α/β

When wild-type mouse embryonic fibroblasts (MEFs) were stimulated with IFN- β , a notable increase in the level of p53 was observed (Fig. 1a). A similar observation was made in cells stimulated with IFN- α (data not shown) but not with IFN- γ (Supplementary Fig. 1). The increase in p53 level with IFN- β treatment was dose-dependent, with about fourfold induction achieved at a high concentration of IFN- β (Fig. 1b). The p53 induction by IFN- β was also observed in two human hepatic cancer cell lines, HepG2 and HLE, p53 function being abrogated in the latter cells by a mutation in the DNA-binding domain¹⁴ (Fig. 1c). We performed a pulse-chase experiment to examine whether the observed p53 protein induction is secondary to suppression of the p53 degradation pathway by IFN- β , typically the MDM2-mediated pathway^{15–17}. However, no difference was observed in the half-life of p53 (40–45 min) between the IFN-treated and untreated MEFs (Fig. 1d), suggesting that p53 protein synthesis is induced by IFN- β stimulation.

Induction of the *p53* gene by IFN- α/β

Information is limited about the induction of the *p53* gene¹⁸, and the above results prompted us to examine whether IFN- α/β induces *p53* gene transcription. Inspection of mouse and human *p53* genes has revealed sequences characteristic of the interferon-stimulated response element (ISRE) within their promoter or first-intron

regions (Fig. 2a); ISRE is activated by the IFN-activated transcription factor ISGF3, a heterotrimeric complex consisting of a signal transducer and activator of transcription factor 1 (Stat1), Stat2 and interferon regulatory factor 9 (IRF-9)^{19–21}. As shown in Fig. 2b (left panel), p53 messenger RNA is induced by IFN- β in wild-type MEFs by about threefold; however, this induction was not observed in MEFs from mice deficient in the *Irf9* gene (IRF-9^{–/–} MEFs). Consistently, p53 protein induction was also abolished in IRF-9^{–/–} MEFs (Fig. 2b, right panel). Similar mRNA induction was observed in IFN- α -stimulated wild-type MEFs (data not shown) as well as IFN- β -stimulated HepG2 cells (Supplementary Fig. 2a). These results suggest that p53 mRNA induction by IFN- β is caused by the transcriptional activation of the gene by ISGF3.

Activation of the *p53*-derived ISREs by ISGF3

Of note, *p53* gene induction by IFN- β is not as strong as that of conventional IFN-inducible genes such as 2'-5' oligoadenylate synthetase (OAS). In MEFs, the induction of p53 and OAS mRNAs are about threefold and eightfold, respectively (Fig. 2b; see below). This difference may be due to the lower affinities of the p53 ISREs (denoted as p53-ISRE1 and p53-ISRE2 in Fig. 2a) for ISGF3. To verify this point, we compared the affinities of the p53-ISREs by electrophoretic mobility shift assay (EMSA) using OAS ISRE as the probe for ISGF3. As shown in Fig. 2c, approximately eightfold higher concentration of the p53-ISREs relative to that of the OAS ISRE is required to inhibit OAS ISRE-ISGF3 complex formation by 50%. Similar observations were made for the two human p53 ISREs (h-p53-ISRE1 and h-p53-ISRE2; Fig. 2a, see also Supplementary Fig. 2b).

Consistently, a transient reporter gene assay revealed that the mouse p53-ISREs are activated by IFN- β , but they need to be multimerized five times to achieve an induction level of the reporter gene similar to that achieved by the monomeric ISRE of the OAS gene (Fig. 2d). To examine ISGF3 binding to the p53 ISRE sequences *in vivo*, we performed chromatin immunoprecipitation (ChIP) assay by treating wild-type MEFs with IFN- β , followed by immunoprecipitation with an anti-Stat2 antibody. As shown in Fig. 2e, DNA

bands diagnostic to p53-ISRE1 and p53-ISRE2 were detected in the cells stimulated with IFN- β . These results *in toto* indicate that IFN- β induces p53 gene transcription through the activation of ISGF3, which binds to the p53 ISREs.

Suppression of oncogene-induced cell transformation

The above results prompted us to examine whether IFN- β induces activation of p53. However, no evidence was observed for the serine phosphorylation of p53 in the IFN- β -stimulated MEFs (see Fig. 4a, left panel), and the target genes of p53 were not induced in these cells (Supplementary Fig. 3). This is perhaps not surprising in view of the fact that IFN- β does not evoke p53 responses in normally growing cells. Hence, the p53 induction by IFN- β may be considered as a means to boost a p53 response to stress signals, suggesting a role for IFN- β induction of p53 in tumour suppression.

To address this issue, we first examined the effect of IFN- β in cells in which p53 is under negative regulation by an oncogene. One of the best-known examples is the E6 gene of the human papilloma viruses (HPV), the product of which targets p53 by inducing its degradation by the ubiquitin proteolytic pathway²². Persistent infection with oncogenic HPV types represents a major risk factor for the development of cervical cancer, in which E6 is known to have

a critical role (together with another oncoprotein, E7 (ref. 23)). When HPV16 E6 is expressed in MEFs, a notable decrease in the basal p53 protein level is observed, but this is countered by IFN- β (Fig. 3a, left panel). Interestingly, the induction of p53 mRNA and protein continues to increase in these cells at least up to 9 h after stimulation (Fig. 3a). This is not due to the IFN action on the E6-mediated degradation of p53 (Supplementary Fig. 4a). The mechanism of this sustained induction is therefore currently unknown but, interestingly, a very similar observation was made with HeLa cells, which are derived from human cervical cancer and express the E6 protein (Supplementary Fig. 4b).

Primary MEFs do not become transformed when HPV E6 is expressed alone; however, they do so when an additional oncogene such as activated Ha-Ras is co-expressed. As shown in Fig. 3b, the numbers of transformed colonies in MEFs expressing E6 and Ha-Ras are markedly suppressed by IFN- β in a dose-dependent manner (about 80% inhibition at a concentration of 10^3 U ml⁻¹). In these cells, p53 mRNA is also IFN- β -inducible, accompanied by enhanced phosphorylation of p53 (Ser 18) (Supplementary Fig. 5a, b). On the other hand, the expression of the mRNAs for E6 and mTERT, a telomerase subunit known to be E6-inducible²⁴, remained unaffected by IFN- β (Supplementary Fig. 6). Of note,

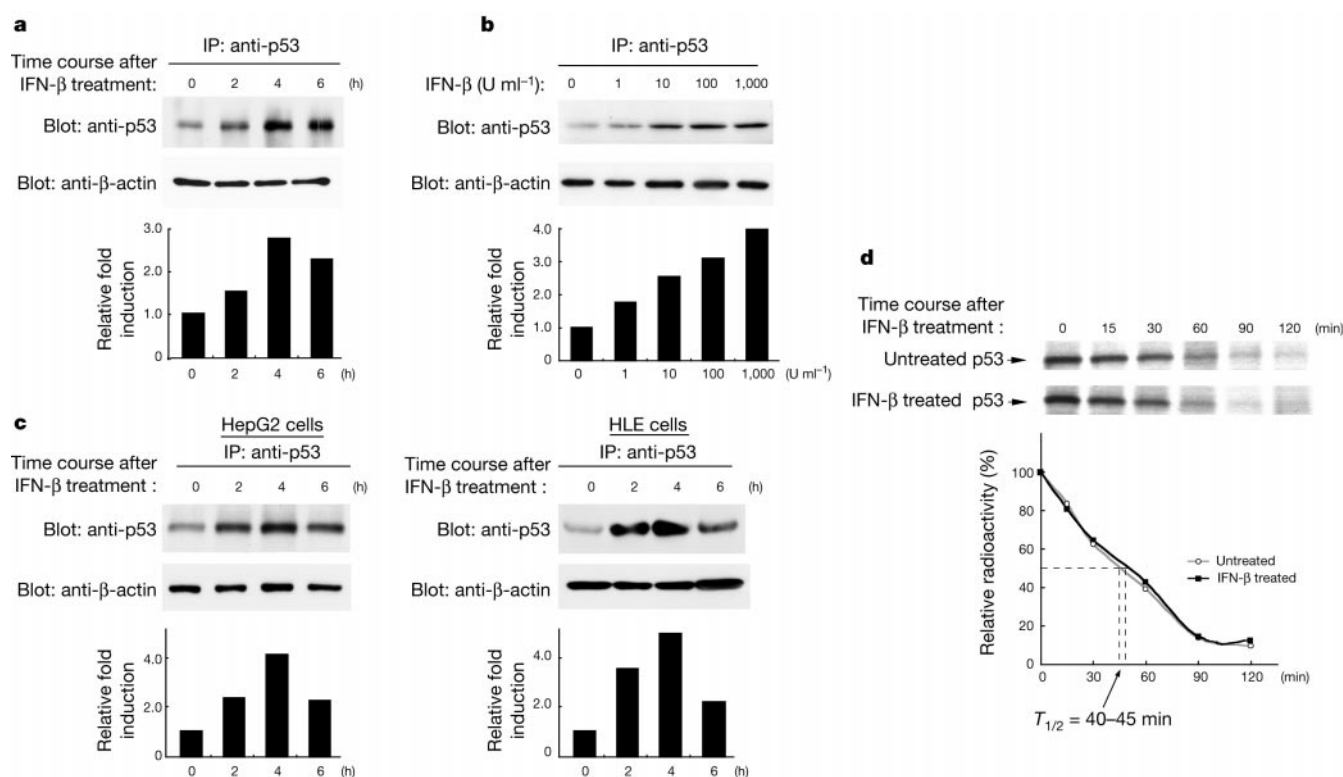


Figure 1 Induction of p53 protein by IFN- β . **a**, MEFs were treated with mouse IFN- β and cell lysates were immunoprecipitated (IP) with anti-p53 antibody. Immunoprecipitates were separated by 9% SDS-PAGE, and immunoblotting analysis (blot) was carried out with the anti-p53 antibody (anti-p53). The supernatants after anti-p53 immunoprecipitation were directly subjected to immunoblotting with the anti- β -actin antibody as the control. We chose this method, as western blotting of the whole-cell lysates gives a high background, but a consistent result was obtained with this approach as well (see Supplementary Fig. 10). The lower panel shows the quantitative display of the induction levels of p53 protein, which were standardized to the intensities of each blot by anti- β -actin, as determined using a densitometer. These and the following experiments were repeated at least three times and the results were highly reproducible.

b, Dose-dependent induction of p53 by IFN- β . MEFs were treated with various

concentrations of IFN- β , and cell lysates were subjected to immunoprecipitation and immunoblot analyses as described above. Results of the quantitative analysis for the induction levels of the p53 protein are shown in the lower panel. **c**, Induction of p53 protein in human hepatocellular carcinoma cell lines. Analysis was performed as described in **a**, except that the cells were stimulated with human IFN- β . **d**, Effect of IFN- β on degradation of p53 protein. After pulse labelling with [³⁵S]methionine and [³⁵S]cysteine, MEFs were treated with or without IFN- β . Cell lysates were prepared at the indicated times and subjected to immunoprecipitation with the anti-p53 antibody and analysed by SDS-PAGE. The incorporated radioactivity was measured by BAS5000 (Fujix) and shown in the lower panel. IFN- β did not affect the degradation of p53 protein in human HepG2 cells (data not shown).

such suppression was not observed in the MEFs from *p53*-deficient mice (*p53*^{-/-} MEFs²⁵), which undergo transformation as induced by Ha-Ras alone (Fig. 3b). Thus, IFN- β exerts a strong anti-oncogenic activity in E6-dependent cell transformation (Fig. 3b), and this is probably mediated by sustaining *p53* expression through transcriptional induction of the gene.

We also examined the effect of IFN- β on the DNA damage-induced apoptotic response of MEFs expressing the adenovirus E1A oncoprotein, as it is well known that this response is *p53*-dependent and critical for tumour suppression²⁶. Binding of annexin V-fluorescein isothiocyanate (FITC; annexin V staining)

and cell viability assay revealed that pretreatment of the MEFs with IFN- β resulted in a notable enhancement of apoptosis in response to X-ray irradiation (Fig. 3c, and data not shown). Congruent with this observation is the finding that *p53* with serine phosphorylation was increased by IFN- β treatment in E1A-expressing MEFs (Supplementary Fig. 7).

Enhancement of cancer cell apoptosis by IFN

The induction of *p53* by IFN- α/β suggests that IFN-treated cells are more susceptible to *p53*-dependent apoptosis in response to DNA-damaging agents such as chemotherapeutic agents used in

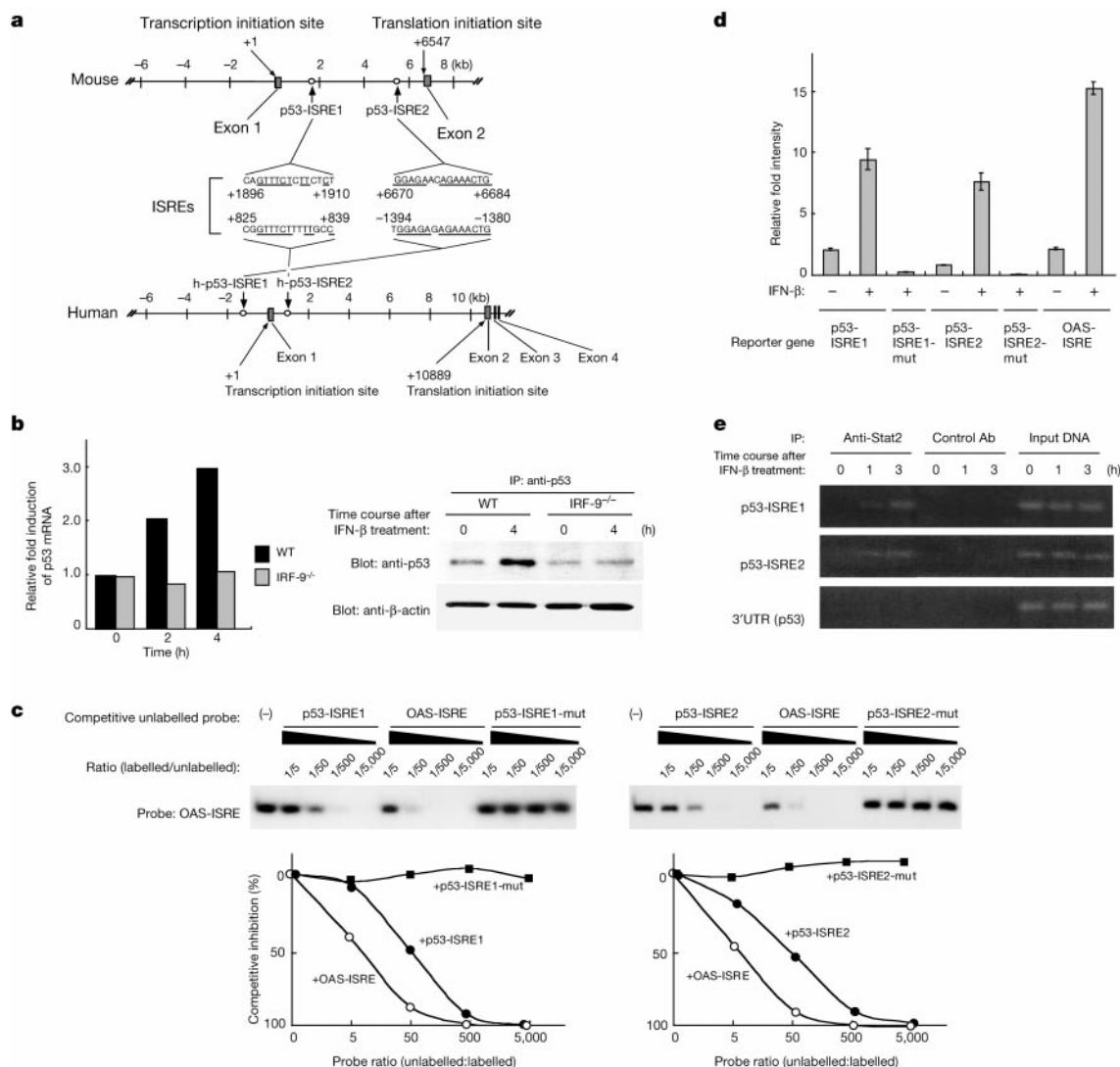


Figure 2 Transcriptional induction of *p53* gene by IFN- β . **a**, Putative ISREs in proximal promoter regions of the murine and human *p53* genes are shown. **b**, Wild-type (WT) or IRF-9^{-/-} MEFs were stimulated with IFN- β for the indicated periods, and the induction of *p53* mRNA levels was analysed by quantitative real-time RT-PCR (left). Four hours after IFN- β treatment, cell lysates were also extracted, and *p53* protein level was evaluated by immunoprecipitation and immunoblotting analysis with the anti-*p53* antibody and anti- β -actin antibody (right). **c**, Lysates from the MEFs stimulated by IFN- β were prepared and subjected to EMSA. Competition assay was carried out with an excessive amount of unlabelled *p53*-ISRE probes. The ratios of the amount of radiolabelled to unlabelled probes are indicated above each lane. OAS-ISRE is an ISRE probe derived from the *OAS* gene. mut-*p53*-ISRE1 or mut-*p53*-ISRE2 is a probe with mutations in the consensus sequence (see Methods). **d**, Transcriptional activation by IFN- β through the *p53*-gene-derived ISREs. After transfection with the luciferase reporter plasmids, *p53*-ISRE1-luc or

p53-ISRE2-luc, MEFs were stimulated with IFN- β , and induction level of luciferase activity with or without IFN- β stimulation was subsequently measured. pOAS-ISRE-luc was used as a positive control for ISGF3-dependent activation. **e**, Chromatin immunoprecipitation assay results. ISGF3 binding to the endogenous ISREs in the *p53* gene was examined using MEFs stimulated with IFN- β . PCR was carried out to detect endogenous ISREs in the gene (*p53*-ISRE1 and *p53*-ISRE2) in immunoprecipitated chromatin fragments: Lanes 1, 2 and 3 show PCR amplification of target sequences in immunoprecipitated chromatin fragments with the anti-Stat2 antibody. Lanes 4, 5 and 6 show the results of PCR using immunoprecipitated samples with anti-control antibody (Ab). Input (lanes 7, 8 and 9) represents PCR amplification of the total input DNA. The results of PCR with primers that detect the 3'-UTR of the *p53* gene are shown in the bottom panel. The results were highly reproducible.

cancer therapy. 5-Fluorouracil (5-FU) is one agent whose efficacy closely correlates with the p53 status of cancer cells²⁷. In view of the induction of p53 by IFN- β in human cancer cell lines HepG2 and HLE (Fig. 1c), we examined whether IFN- β enhances the apoptotic response in these cells. As shown in Fig. 3d, loss of viability of HepG2 cells was significantly enhanced by IFN- β in a dose-dependent manner at a concentration of 5-FU that itself has only a slight effect on the cell. In contrast, such an effect was not seen in HLE cells, in which p53 is functionally inactive¹⁴. These results can be interpreted to mean that the IFN- β -induced p53 apoptotic response comes into action in these cells; a severalfold increase in p53 level would make

HepG2 cells sensitive enough to respond to sub-optimal doses of 5-FU.

It is known that HeLa cells undergo apoptosis in response to DNA damage in a p53-dependent manner^{28,29}. In view of our finding that p53 is also induced by IFN- β in this cell line, we asked whether the apoptotic response is enhanced by IFN- β . As shown in Fig. 3e, significant enhancement was observed in X-ray irradiation-induced apoptosis.

Activation of p53 in virally infected cells

It has been reported that p53 function and IFN signalling are

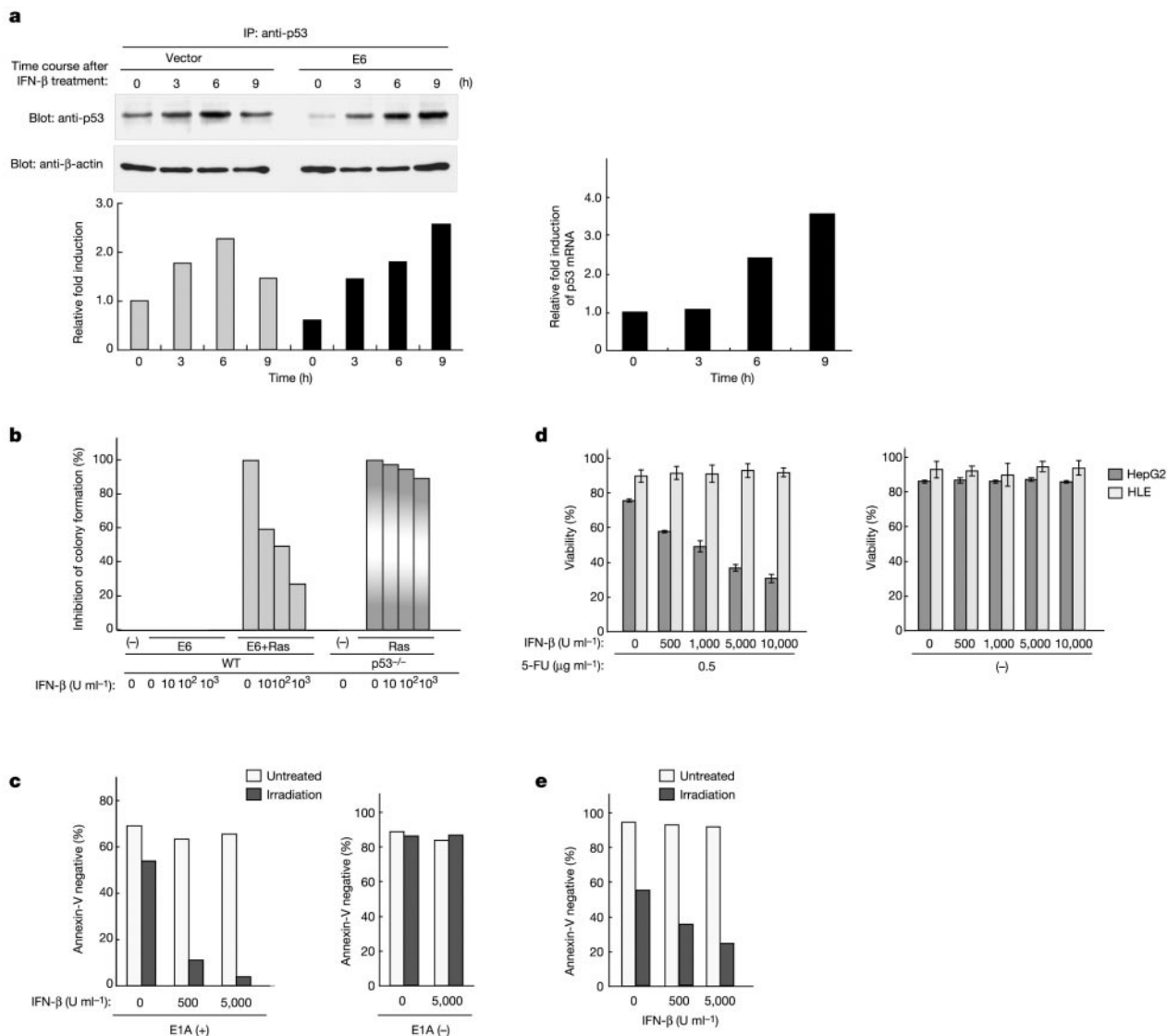


Figure 3 Suppression of oncogene-induced cell transformation and enhancement of oncogene-dependent apoptosis by IFN- β . **a**, Sustained upregulation of p53 protein level by IFN- β in MEFs expressing type 16 HPV-derived E6 oncoprotein. MEFs were transfected with pLRHL (vector) or pLRHL-16E6 (E6) via retroviral gene transfer, and after IFN- β treatment the induction level of the p53 protein or mRNA at each time point was analysed by immunoprecipitation and immunoblotting analysis (left) or quantitative real-time RT-PCR (right). **b**, Suppression of oncogene-induced colony formation by IFN- β . Wild-type MEFs, retrovirally transduced with only HPV16 E6 protein or both E6 and activated Ha-Ras, or p53^{-/-} MEFs expressing activated Ha-Ras, were seeded on soft agar in 60-mm dishes in the absence or presence of IFN- β , and were incubated for three weeks. The medium with the same IFN- β concentration was changed every three days. The histogram represents the relative number of colonies (%) compared with that of colonies

cultured in the absence of IFN- β . Results are a representative of two independent experiments. **c**, Enhancement of DNA-damage-induced apoptosis in E1A-expressing MEFs after pretreatment with IFN- β (left). The percentage of annexin V-FITC-bound cells was determined 40 h after X-ray irradiation. Radiation-induced apoptosis and its enhancement by IFN- β were not observed in MEFs that do not express E1A (right). **d**, HepG2 or HLE cells were incubated with 5-FU (0.5 μ g ml⁻¹) alone or in combination with IFN- β . All cells were collected 72 h after the addition of 5-FU, and viable cells were counted by trypan blue exclusion staining. The proportion of cells incubated without 5-FU was defined as 100% viability. **e**, Enhancement of radiation-induced apoptosis in HeLa cells after the pretreatment with IFN- β . The percentage of annexin V-FITC-bound cells was determined 40 h after X-ray irradiation.

inhibited by products of oncogenic viruses^{23,30–34}; however, it is unknown whether p53 itself has any role in the antiviral defence mechanism. Previous studies have shown that *p53*-deficient mice show no sign of abnormality in T and B cells, and that they can mount adaptive immune responses normally³⁵. In view of the newly identified connection between p53 and IFN- α/β , we investigated whether p53 has a role in antiviral responses.

When MEFs were infected with vesicular stomatitis virus (VSV), marked phosphorylation of p53 was detected by an antibody against Ser 18 (Fig. 4a, left panel). This p53 phosphorylation was also observed in MEFs infected with other types of virus, such as Newcastle disease virus (NDV) and herpes simplex virus (HSV) (Fig. 4a, right panel). Furthermore, the same observation—that is, phosphorylation of Ser 15 of human p53 (ref. 36)—was made in the VSV-infected HepG2 cells (Fig. 4b).

The ATM kinase is known to be involved in this phosphorylation in response to DNA damage^{37,38}. We therefore asked whether the virus-induced phosphorylation of Ser 18 is also ATM-dependent in the following experimental setting. *p53*-deficient and *p53/Atm* doubly deficient MEFs³⁹ were infected with a p53-expressing retrovirus, and the phosphorylation of the expressed p53 in the presence or absence of ATM was examined after infection by VSV or HSV. Perhaps surprisingly, as shown in Fig. 4c, phosphorylation of p53 at Ser 18 by VSV or HSV was hardly detectable in MEFs deficient in the

Atm gene, suggesting an integral role of ATM in the virus-induced phosphorylation of p53. We infer that, similar to DNA damage, ATM activation is caused by changes in the chromatin structure in virus-infected cells⁴⁰.

Induction of p53 target genes by virus

It was difficult to examine critically the serine phosphorylation at other sites, because the available antibodies did not react well even with the p53 from irradiated MEFs. Therefore, we examined the induction of p53 target genes in MEFs after VSV infection, to test whether virus infection results in p53 activation. RNA blotting analysis revealed that mRNAs for *Mdm2* and *Puma* are clearly inducible by VSV: the mRNA induction begins 8 h and peaks 12 h after infection (Fig. 4d, left panel). Interestingly, mRNAs for *p21*^{WAF1/Cip1} and *Noxa*, which are both induced in response to DNA damage (Fig. 4d, right panel; see also ref. 41) are not induced by the virus. These results suggest that p53 is indeed in its active form in VSV-infected cells, but that its activation profile of target genes is distinct from that induced by DNA-damaging agents. Although the mechanism of this differential activation of p53-inducible genes is currently unknown, it is interesting that *p21*^{WAF1/Cip1}, the induction of which acts in favour of cell survival (ref. 42 and T. S., unpublished data), is not virus-inducible—the absence of *p21*^{WAF1/Cip1} induction would favour p53-mediated

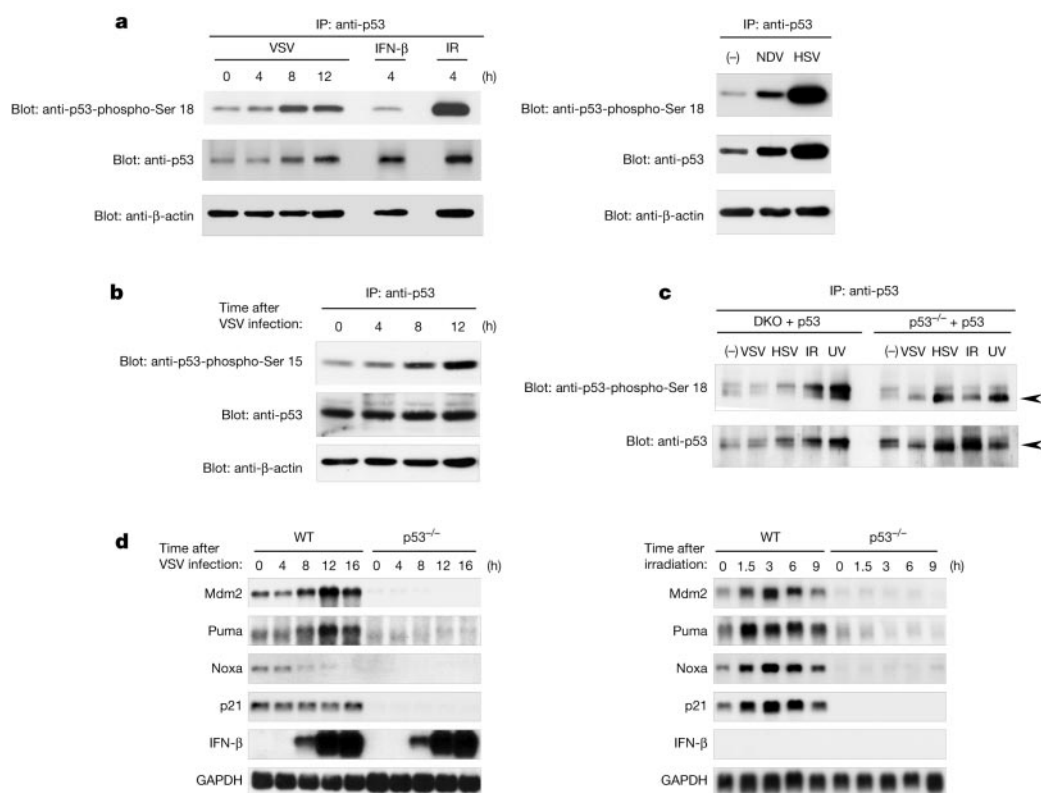


Figure 4 Activation of p53 by viruses. **a**, Induction of phosphorylation of p53 at Ser 18 by virus infection. MEFs were infected with VSV (left), NDV or HSV (right), or treated with IFN- β for 4 h, and cell lysates were analysed by immunoprecipitation with the anti-p53 antibody and immunoblotting with specific antibodies against phospho-Ser 18 or p53. Phosphorylation of Ser 18 at 4 h after X-ray irradiation (IR) is shown as a positive control. Note that p53 phosphorylation is not induced by IFN- β . **b**, VSV-induced phosphorylation of human p53 at Ser 15. Human HepG2 cells were infected by VSV and phosphorylation of Ser 15 was examined as described above for mouse p53. **c**, ATM-dependent phosphorylation of p53-Ser 18 by virus infection. p53^{-/-} MEFs and MEFs doubly deficient for p53 and ATM (DKO) were infected with p53-expressing retrovirus, and the

phosphorylation status of Ser 18 was examined 8 h after VSV infection or ultraviolet irradiation (UV), or 4 h after X-ray irradiation (IR), by immunoprecipitation and immunoblotting analysis as described in **a**. The arrow indicates the band corresponding to p53 protein, whereas an additional band above this arrow represents a nonspecific band that is also detected in the MEFs transfected with a control vector (data not shown). It is known that p53 phosphorylation by UV is dependent on another kinase, ATR. **d**, Differential activation of p53 target genes after viral infection or X-ray irradiation. After VSV infection for the indicated time periods, the kinetics of p53-dependent gene induction were analysed by RNA blotting in wild-type or p53^{-/-} MEFs. The result of the induction of target genes of p53 after X-ray irradiation is shown in the right panel.

apoptosis. It is possible that this difference is a reflection of the status of p53 phosphorylation and/or activation of other transcription factors acting cooperatively with p53 on these target genes.

p53-dependent apoptosis in virus-infected cells

Activation of p53 and induction of its proapoptotic target genes in virally infected cells suggest that these events lead to an apoptotic response of the infected cells, an event that can be considered as altruistic suicide that limits virus replication. In fact, as shown in Fig. 5a, wild-type MEFs infected with VSV eventually underwent apoptosis as revealed by annexin V staining, but this response was strongly suppressed in p53^{-/-} MEFs. In this regard, it is worth noting that the upregulation of p53 in response to VSV infection can be detected in wild-type MEFs but not in IFN- α/β receptor 1 (IFNAR1)-deficient MEFs (IFNAR1^{-/-} MEFs) (Fig. 5b). It is worth highlighting two features of p53 expression and phosphorylation. First, p53 phosphorylation (at Ser 18) is induced in IFNAR1^{-/-} MEFs, indicating that the IFN signal is not necessary for this event. In fact, this observation is consistent with our notion that IFN does not activate p53 but contributes to the enhancement of the p53 response by inducing the p53 gene. Second, the p53 protein level is slightly increased by VSV infection in IFNAR1^{-/-} MEFs, suggesting that VSV may stabilize the p53 protein. This may be an interesting issue to be addressed further. Regarding the contribution of IFN-induced p53 in the virus-induced apoptotic response, we also found that IFNAR1^{-/-} MEFs are indeed more resistant to VSV-induced apoptosis than wild-type MEFs (Supplementary Fig. 8). Therefore, although careful interpretation of

this observation will be required, it is possible that p53 induction by virus-induced IFN- α/β may at least in part enhance the apoptotic response in virally infected MEFs.

To determine whether the absence of p53 response affects the extent of virus replication, we used the following assay. Wild-type MEFs were infected with VSV as described above, and virus titres were monitored in the supernatant of the infected cells after they all died as a result of the infection. As shown in Fig. 5c, the virus yield was more than 30-fold higher in p53^{-/-} MEFs than in wild-type MEFs. These results suggest that the p53-dependent apoptosis in virally infected cells contributes to limiting virus replication. In order to verify the role of p53 in antiviral response *in vivo*, we examined the susceptibility of p53^{-/-} mice to VSV. Surprisingly, as shown in Fig. 5d, these mice succumbed to VSV infection. In addition, a marked increase (about 100-fold) was observed in the sera of p53^{-/-} mice compared with wild-type mice (Fig. 5e). These observations lend further support to the notion that the p53 response to virus infection constitutes a critical aspect of antiviral immunity.

Discussion

The regulation of the tumour suppressor p53 has been extensively studied at the protein level; however, little is known about the transcriptional regulation of the p53 gene. In this study, we have shown that the p53 gene is transcriptionally induced by IFN- α/β through ISGF3 activation, demonstrating p53 gene induction by a cytokine. Whereas IFN- α/β induces p53 mRNA and increases its protein level, p53-mediated responses such as cell cycle arrest or

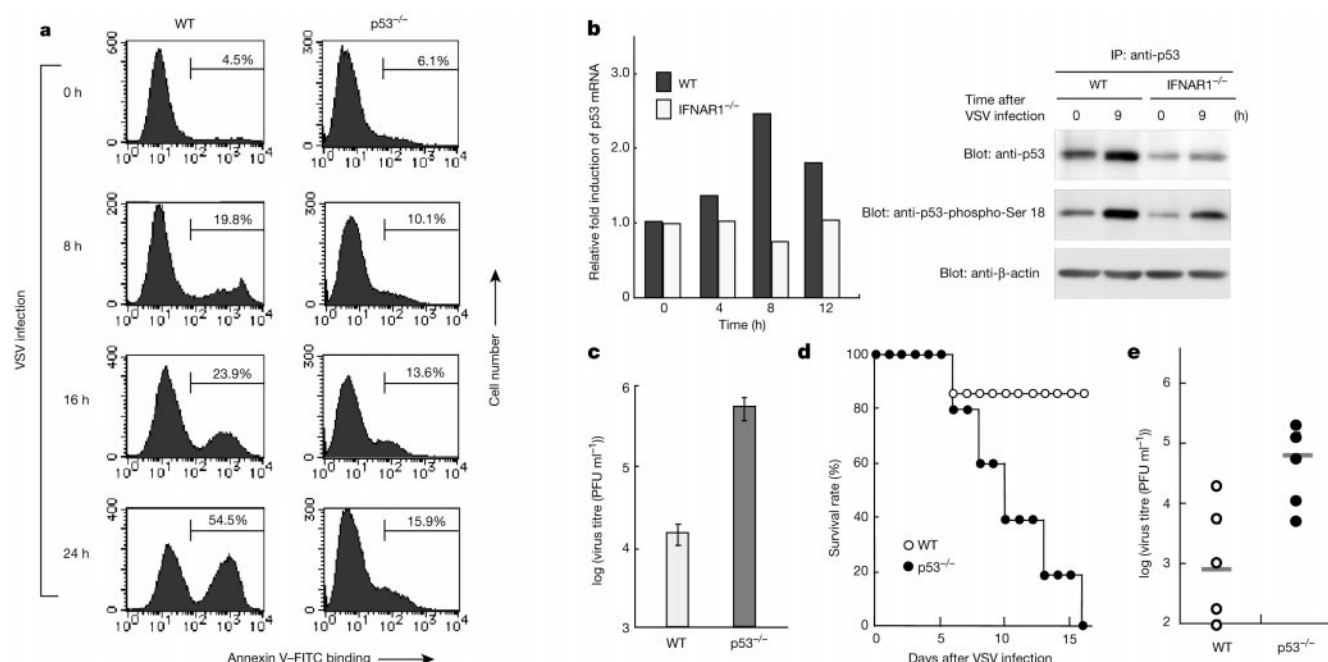


Figure 5 Role of p53 in antiviral defence. **a**, p53-dependent apoptosis in virally infected cells. Wild-type and p53^{-/-} MEFs were infected by VSV, and the apoptotic response was analysed by annexin V-FITC staining. **b**, Upregulation of the levels of p53 mRNA after VSV infection in wild-type MEFs but not in IFNAR1^{-/-} MEFs (left). The induction levels of p53 mRNA at the indicated times after VSV infection were analysed by quantitative real-time RT-PCR analysis, and normalized to the expression levels of the 28S rRNA. The upregulation of p53 protein level in wild-type and IFNAR1^{-/-} MEFs in response to VSV infection is shown (right). Wild-type and IFNAR1^{-/-} MEFs were infected with VSV, and at 9 h after infection cell lysates were immunoprecipitated (IP) with the anti-p53 antibody. Subsequently, immunoblotting analysis (blot) was carried out with the anti-p53 or anti-p53-phospho-Ser 18 antibody. **c**, Virus titres of the supernatant of the wild-type and

p53^{-/-} MEFs after the viability of all the cells was lost after VSV infection (10 M.O.I.). **d**, Survival rate of wild-type and p53^{-/-} mice after VSV infection. Ten wild-type and nine p53^{-/-} mice were intravenously infected with VSV (2×10^6 plaque-forming units (PFU)), and the survival of these mice was monitored daily. All of the p53^{-/-} mice died within 16 days after infection, whereas the wild-type mice, except for one, survived. **e**, Virus titres in sera of virus-infected mice. Wild-type and p53^{-/-} mice were infected with VSV as described in **d**. Sera from each mouse were collected two days after infection, and VSV titres in the sera were determined by the plaque-formation assay. The bars indicate the mean values of the virus titre. The difference in virus titre between wild-type and p53^{-/-} mice was statistically significant ($P < 0.02$, Student's *t*-test).

apoptosis were not observed in cells treated with IFN- α/β alone. Although seemingly futile, this induction could therefore be viewed as a regulated 'trade-off' to provide the cell with a greater dynamic range in its response to a variety of stress signals. This notion is supported by the present findings that cells treated with IFN- β evoke more robust p53 responses than do untreated cells (Fig. 3). Notably, induction of the p53 mRNA by IFN- β is relatively weak compared with that of the mRNAs of conventional IFN-inducible genes. It is likely that the low affinity of the p53-ISREs to IFN-activated ISGF3 accounts for the weak induction (Fig. 2). In view of the fact that p53 is an extremely efficient inhibitor of cell growth, the modest induction of p53 by IFNs may help to avoid adverse p53 actions in normal growing cells.

There are numerous reports on the usefulness of IFN- α/β for the treatment of some types of human cancer, including HPV-associated cervical cancer and hepatic cancer^{43–45}. However, the molecular basis of IFN action in cancer treatment still remains largely elusive. Our study suggests that one mechanism of the anti-tumour action of IFN may involve p53 induction. In this regard, it is interesting that IFNAR1^{-/-} MEFs undergo spontaneous transformation *in vitro* and that the mutant mice develop papilloma of the skin at a high incidence (about eightfold relative to wild-type mice) when treated with chemical carcinogens (A.T. and N. Tanaka, unpublished data). In addition, our study suggests the possible usefulness of treating human cancers with IFN- α/β in combination with chemotherapeutic drugs that activate p53. Combined therapy with IFN- α/β and chemotherapeutic drugs such as 5-FU may permit the use of lower doses of chemotherapeutic agents, which would otherwise exert toxic side effects by p53-independent mechanisms. This will be a clinically interesting issue to address further.

Our results reveal an important role of p53 in the complex innate antiviral host defences (Fig. 5). Prompt induction of apoptosis of virus-infected cells via p53 activation will be beneficial to the host in limiting virus replication: virus production would be suppressed if the infected cells undergo rapid apoptosis before the onset of cellular lysis (Fig. 5a). Apoptotic cells would be swiftly engulfed by phagocytic cells *in vivo* and this event would also contribute to inhibiting the spread of virus (Fig. 5e). This p53 response is not absolutely dependent on but could be enhanced by IFN- α/β . Interestingly, in the absence of p53, IFN-treated cells are more resistant to apoptosis on infection with VSV or encephalomyocarditis virus (EMCV) (Supplementary Fig. 9a, b), and the total VSV virus yield becomes higher in p53^{-/-} MEFs than in wild-type MEFs when these cells were treated by IFN- β followed by the infection (Supplementary Fig. 9c). On the basis of these findings, one may infer the following events: at the early phase of virus infection, virus-infected cells produce IFN- α/β and eventually undergo p53-dependent apoptosis. On the other hand, virus-induced IFN- α/β may act on the surrounding, uninfected cells to help antiviral defences by inducing cellular genes that inhibit virus replication and, in addition, by inducing p53 to prime cells for enhanced apoptosis. However, the details of how IFN- α/β and p53 cooperate in antiviral immunity need to be clarified further. Our study reveals a hitherto unrecognized cooperation between p53 and IFN- α/β , providing a new link between tumour suppression and antiviral host defence. □

Methods

Mice, cell cultures, viruses and reagents

The generation of IRF-9^{-/-} and p53^{-/-} mice has been described previously^{25,46}. IFNAR1^{-/-} mice were purchased from B&K Universal Group Ltd. MEFs were prepared following a standard procedure⁴⁷. The above mice were all on the C57BL/6 background. ATM and p53-double-deficient MEFs³⁹ were provided by C. Westphal. The human hepatoma cell lines HepG2 and HLE were purchased from the Health Science Research Resources Bank. Recombinant human and mouse IFN- β were provided by Toray Co., Ltd. Unless otherwise stated, the concentration of IFN- β used in the assays was 500 U ml⁻¹. The infection of MEFs with VSV, HSV, NDV and EMCV was carried out as described

previously⁴⁸, and multiplicity of infection (M.O.I.) is 1.0 unless stated otherwise. X-ray irradiation was performed at 30 Gy.

Immunoprecipitation and immunoblotting

Cell lysis, immunoprecipitation and immunoblotting were carried out as described⁴⁷. Antibodies against the following proteins were purchased: p53 (Ab-1; Oncogene Research Products), p53 (FL-393; Santa Cruz Biotechnology) and β -actin (AC-15; Sigma-Aldrich). For the detection of mouse p53 phosphorylation on Ser 18, phospho-Ser 15-specific antiserum was provided by Y. Taya³⁶.

RNA analysis

RNA extraction and polymerase chain reaction with reverse transcription (RT-PCR) analysis were performed as described previously⁴⁷. Quantitative real-time RT-PCR was performed with the Lightcycler and SYBR Green system (Roche Molecular Biochemicals), and the data were normalized by the expression level of 28S ribosomal RNA for each sample. The following oligonucleotide primers specific to mouse p53 and 28S rRNA were used: p53, 5'-ACTGCATGGACGATCTGTTG-3' (sense) and 5'-GCCATAGTTGCCCTGGTAAG-3' (antisense); 28S rRNA, 5'-CAGGGGAATCCGACTGTTTA-3' (sense) and 5'-ATGACGAGGCATTGGCTAC-3' (antisense). For RNA blot analysis, equal amounts of total RNA (5 μ g) were loaded in each lane. Probes for Mdm2, Noxa, p21^{WAF1/Cip1}, IFN- β and GAPDH mRNAs have been described previously^{41,48}. A hybridization probe for Puma mRNA was generated by PCR amplification of a complementary DNA fragment corresponding to codons 262–843 of the Puma gene.

Retroviral expression

The expression of activated Ha-Ras was mediated by retroviral gene transfer, wherein the pGDV12ras was used as described previously⁴⁹. The details of the retroviral expression vectors for the HPV16 E6 protein, E1A and wild-type p53 are mentioned in Supplementary Information. Retroviral transduction into MEFs was carried out as described previously^{47,48}.

Apoptosis assay

Annexin V staining with FITC-conjugated anti-annexin V (Molecular Probes) was carried out, and samples were subjected to flow cytometric analysis with FACS calibur (Becton & Dickinson).

Chromatin immunoprecipitation assay

The ChIP assay was performed according to the manufacturer's protocol (Upstate Biotechnology) with a slight modification. Briefly, MEFs were treated with or without IFN- β for 1.5 and 3.0 h before formaldehyde cross-linking. The specific antibodies used for immunoprecipitations were the anti-Stat2 antibody (Upstate Biotechnology) and a control antibody (rabbit IgG). After protein–DNA cross-links in the immunoprecipitates were reversed, the purified DNA was analysed by PCR (35 cycles; 30 s at 95 °C, 60 s at 55 °C, 60 s at 72 °C) with primers that detect sequences containing p53-ISRE1 (nucleotide number +1757 to +1901) or p53-ISRE2 (+5619 to +5780), and a pair of negative-control primers that recognize the DNA sequence (+11093 to +11250) of the 3' untranslated region (UTR) of the p53 gene. The PCR products were visualized on an ethidium bromide gel.

Transformation assay

Anchorage-independent colony formation was carried out as described previously⁴⁹. Cells were seeded at 1×10^5 cells into 60-mm dishes in a suspension of 1.3% methylcellulose gel dissolved in DMEM medium supplemented with 10% FCS on top of a bed composed of 0.53% agarose in the same culture medium, and incubated with or without the indicated amounts of IFN- β . The number of colonies formed was determined three weeks after the seeding.

Measurement of virus titre

Virus titre in the cells or mice infected with VSV was determined as described previously⁴⁶.

EMSA

We analysed Celera proprietary human and murine genomic databases of the p53 gene and identified two putative ISREs upstream of the transcriptional initiation site and in the first intron using the TRANSFAC database. For EMSA, competition assay was carried out with different concentrations of each non-radiolabelled oligonucleotide probe containing the element of p53-ISRE1 or p53-ISRE2, as well as those of mut-p53-ISRE1 and mut-p53-ISRE2, which have mutated sequences in the putative ISREs: p53-ISRE1, +1891 CAGTTTCTCTCTCTCT +1915; p53-ISRE2, +5665 GGAGAACAGAACTG +5689; mut-p53-ISRE1, CAGCCATTCCTCTCT; mut-p53-ISRE2, GGAGAGCAATGGCTG. After MEFs were treated with IFN- β for 45 min, EMSA was performed as described previously⁴⁷. Equal amounts of proteins from whole-cell extracts were loaded onto each lane.

Assays of transcriptional activity

Oligonucleotides corresponding to region +1892 to +1913 or region +5668 to +5689 of the murine p53 gene were synthesized. Each of the five copies of the oligonucleotide was tandemly multimerized and ligated to a fragment (region –55 to 0 of the IFN- β gene (*Irfb*) promoter containing a TATA box and a cap site (ref. 50)). The resultant fragments were cloned into the *SacI* and *KpnI* sites of the pGL2-basic vector (Promega) (p53-ISRE1-luc and p53-ISRE2-luc, respectively). Similarly, mut-p53-ISRE1-luc and mut-p53-ISRE2-luc were constructed by introducing the same mutations in the ISRE sequence as described above. After transfection, MEFs were stimulated with IFN- β for 8 h and luciferase activities were analysed by the dual luciferase assay (Roche Diagnostics) including the

measurement of β -galactosidase activities that were used as an internal control for transfection efficiency. Twenty-four hours after transfection, cells were treated with IFN- β for 8 h, and cell extracts were prepared and subjected to the luciferase assay. pOAS-ISRE-luc, which contains the ISRE (GGGAAATGGAAACT) from the OAS gene⁴⁶, was used as a positive control. The averages and standard deviations of the values of luciferase activities from the triplicates of a representative experiment are shown.

Received 19 March; accepted 28 May 2003; doi:10.1038/nature01850.
Published online 16 July 2003.

1. Oren, M. Relationship of p53 to the control of apoptotic cell death. *Semin. Cancer Biol.* **5**, 221–227 (1994).
2. Jacks, T. & Weinberg, R. A. Cell-cycle control and its watchman. *Nature* **381**, 643–644 (1996).
3. Prives, C. & Hall, P. A. The p53 pathway. *J. Pathol.* **187**, 112–126 (1999).
4. Vogelstein, B., Lane, D. & Levine, A. J. Surfing the p53 network. *Nature* **408**, 307–310 (2000).
5. Vousden, K. H. & Lu, X. Live or let die: the cell's response to p53. *Nature Rev. Cancer* **2**, 594–604 (2002).
6. Pestka, S., Langer, J. A., Zoon, K. C. & Samuel, C. E. Interferons and their actions. *Annu. Rev. Biochem.* **56**, 727–777 (1987).
7. De Maeyer, E. & De Maeyer-Guignard, J. *Interferons and Other Regulatory Cytokines* (John Wiley and Sons, New York, 1988).
8. Vilcek, J. & Sen, G. S. in *Fields Virology* 3rd edn (eds Fields, D. M., Knipe, P. M. & Howley, P. M.) 375–399 (Lippincott-Raven, Philadelphia, 1996).
9. Biron, C. A. Interferons α and β as immune regulators—a new look. *Immunity* **14**, 661–664 (2001).
10. Gresser, I. & Belardelli, F. Endogenous type I interferons as a defense against tumors. *Cytokine Growth Factor Rev.* **13**, 111–118 (2002).
11. Strander, H. & Einhorn, S. Interferons and the tumor cell. *Biotherapy* **8**, 213–218 (1996).
12. Guterman, J. U. Cytokine therapeutics: lessons from interferon α . *Proc. Natl Acad. Sci. USA* **91**, 1198–1205 (1994).
13. Pfeiffer, L. M. *et al.* Biological properties of recombinant α -interferons: 40th anniversary of the discovery of interferons. *Cancer Res.* **58**, 2489–2499 (1998).
14. Hsu, I. C. *et al.* p53 gene mutation and integrated hepatitis B viral DNA sequences in human liver cancer cell lines. *Carcinogenesis* **14**, 987–992 (1993).
15. Haupt, Y., Maya, R., Kazaz, A. & Oren, M. Mdm2 promotes the rapid degradation of p53. *Nature* **387**, 296–299 (1997).
16. Kubbutat, M. H., Jones, S. N. & Vousden, K. H. Regulation of p53 stability by Mdm2. *Nature* **387**, 299–303 (1997).
17. Sherr, C. J. & Weber, J. D. The ARF/p53 pathway. *Curr. Opin. Genet. Dev.* **10**, 94–99 (2000).
18. Ginsberg, D., Oren, M., Yaniv, M. & Piette, J. Protein-binding elements in the promoter region of the mouse p53 gene. *Oncogene* **5**, 1285–1290 (1990).
19. Darnell, J. E. Jr, Kerr, I. M. & Stark, G. R. Jak-STAT pathways and transcriptional activation in response to IFNs and other extracellular signaling proteins. *Science* **264**, 1415–1421 (1994).
20. Haque, S. J. & Williams, B. R. Identification and characterization of an interferon (IFN)-stimulated response element-IFN-stimulated gene factor 3-independent signaling pathway for IFN- α . *J. Biol. Chem.* **269**, 19523–19529 (1994).
21. Bluyssen, A. R., Durbin, J. E. & Levy, D. E. ISGF3 γ p48, a specificity switch for interferon activated transcription factors. *Cytokine Growth Factor Rev.* **7**, 11–17 (1996).
22. Scheffner, M., Huibregtse, J. M., Vierstra, R. D. & Howley, P. M. The HPV-16 E6 and E6-AP complex functions as a ubiquitin-protein ligase in the ubiquitination of p53. *Cell* **75**, 495–505 (1993).
23. zur Hausen, H. Papillomaviruses and cancer: from basic studies to clinical application. *Nature Rev. Cancer* **2**, 342–350 (2002).
24. Veldman, T., Horikawa, I., Barrett, J. C. & Schlegel, R. Transcriptional activation of the telomerase hTERT gene by human papillomavirus type 16 E6 oncoprotein. *J. Virol.* **75**, 4467–4472 (2001).
25. Tsukada, T. *et al.* Enhanced proliferative potential in culture of cells from p53-deficient mice. *Oncogene* **8**, 3313–3322 (1993).
26. Lowe, S. W., Ruley, H. E., Jacks, T. & Housman, D. E. p53-dependent apoptosis modulates the cytotoxicity of anticancer agents. *Cell* **74**, 957–967 (1993).
27. Bunz, F. *et al.* Disruption of p53 in human cancer cells alters the responses to therapeutic agents. *J. Clin. Invest.* **104**, 263–269 (1999).
28. Butz, K. *et al.* Functional p53 protein in human papillomavirus-positive cancer cells. *Oncogene* **10**, 927–936 (1995).
29. Mendonca, M. S. *et al.* Delayed apoptotic responses associated with radiation-induced neoplastic transformation of human hybrid cells. *Cancer Res.* **59**, 3972–3979 (1999).
30. Neil, J. C., Cameron, E. R. & Baxter, E. W. p53 and tumour viruses: catching the guardian off-guard. *Trends Microbiol.* **5**, 115–120 (1997).

31. Evan, G. & Littlewood, T. A matter of life and cell death. *Science* **281**, 1317–1322 (1998).
32. Levy, D. E. & Garcia-Sastre, A. The virus battles: IFN induction of the antiviral state and mechanisms of viral evasion. *Cytokine Growth Factor Rev.* **12**, 143–156 (2001).
33. Goodbourn, S., Didcock, L. & Randall, R. E. Interferons: cell signalling, immune modulation, antiviral response and virus countermeasures. *J. Gen. Virol.* **81**, 2341–2364 (2000).
34. Katze, M. G., He, Y. & Gale, M. Jr Viruses and interferon: a fight for supremacy. *Nature Rev. Immunol.* **2**, 675–687 (2002).
35. Donehower, L. A. *et al.* Mice deficient for p53 are developmentally normal but susceptible to spontaneous tumours. *Nature* **356**, 215–221 (1992).
36. Shieh, S. Y., Ikeda, M., Taya, Y. & Prives, C. DNA damage-induced phosphorylation of p53 alleviates inhibition by MDM2. *Cell* **91**, 325–334 (1997).
37. Banin, S. *et al.* Enhanced phosphorylation of p53 by ATM in response to DNA damage. *Science* **281**, 1674–1677 (1998).
38. Canman, C. E. *et al.* Activation of the ATM kinase by ionizing radiation and phosphorylation of p53. *Science* **281**, 1677–1679 (1998).
39. Westphal, C. H. *et al.* Genetic interactions between ATM and p53 influence cellular proliferation and irradiation-induced cell cycle checkpoints. *Cancer Res.* **57**, 1664–1667 (1997).
40. Bakkenist, C. J. & Kastan, M. B. DNA damage activates ATM through intermolecular autophosphorylation and dimer dissociation. *Nature* **421**, 499–506 (2003).
41. Oda, E. *et al.* Noxa, a BH3-only member of the Bcl-2 family and candidate mediator of p53-induced apoptosis. *Science* **288**, 1053–1058 (2000).
42. Polyak, K., Waldman, T., He, T. C., Kinzler, K. W. & Vogelstein, B. Genetic determinants of p53-induced apoptosis and growth arrest. *Genes Dev.* **10**, 1945–1952 (1996).
43. Koromilas, A. E., Li, S. & Matlashewski, G. Control of interferon signaling in human papillomavirus infection. *Cytokine Growth Factor Rev.* **12**, 157–170 (2001).
44. Wadler, S. & Schwartz, E. L. Antineoplastic activity of the combination of interferon and cytotoxic agents against experimental and human malignancies: a review. *Cancer Res.* **50**, 3473–3486 (1990).
45. Miyamoto, A. *et al.* Advanced hepatocellular carcinoma with distant metastases, successfully treated by a combination therapy of α -interferon and oral tegafur/uracil. *J. Gastroenterol. Hepatol.* **15**, 1447–1451 (2000).
46. Kimura, T. *et al.* Essential and non-redundant roles of p48 (ISGF3 γ) and IRF-1 in both type I and type II interferon responses, as revealed by gene targeting studies. *Genes Cells* **1**, 115–124 (1996).
47. Takaoka, A. *et al.* Cross talk between interferon- γ and - α/β signaling components in caveolar membrane domains. *Science* **288**, 2357–2360 (2000).
48. Sato, M. *et al.* Distinct and essential roles of transcription factors IRF-3 and IRF-7 in response to viruses for IFN- α/β gene induction. *Immunity* **13**, 539–548 (2000).
49. Tanaka, N. *et al.* Cellular commitment to oncogene-induced transformation or apoptosis is dependent on the transcription factor IRF-1. *Cell* **77**, 829–839 (1994).
50. Fujita, T., Ohno, S., Yasumitsu, H. & Taniguchi, T. Delimitation and properties of DNA sequences required for the regulated expression of human *interferon- β* gene. *Cell* **41**, 489–496 (1985).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank Y. Taya for anti-p53-P-Ser 15 antibody; Y. Hirabayashi and T. Inoue for p53^{-/-} mice; T. Kiyono for pLRHL-16E6; C. Westphal for ATM/p53 deficient MEFs; K. Segawa for pSVp53c17 and pLSVNC-9; T. Nakajima for adenovirus 2 E1A 12S cDNA; Toray Co., Ltd for recombinant human and mouse IFN- β ; J. Vilcek and E. Barsoumian for critical reading of the manuscript; H. zur Hausen, F. Melchers, Y. Taya and N. Tanaka for advice; and T. Yokochi for technical assistance. This work was supported in part by a grant for Advanced Research on Cancer from the Ministry of Education, Culture, Sports, Science, and Technology of Japan, Grants-in-Aid from the Ministry of Education, Science and Culture (Scientific Research on Priority Areas), grants of the Virtual Research Institute of Ageing of Nippon Boehringer Ingelheim, the Welfare Medicinal Research Foundation, Kanai Foundation for Life & Socio-Medical Science, and Senri Life Science Foundation. D.S. was a postdoctoral fellow, supported by the Japan Society for the Promotion of Science.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.T. (tada@m.u-tokyo.ac.jp).

The break-up of heavy electrons at a quantum critical point

J. Custers¹, P. Gegenwart¹, H. Wilhelm¹, K. Neumaier², Y. Tokiwa¹, O. Trovarelli¹, C. Geibel¹, F. Steglich¹, C. Pépin³ & P. Coleman⁴

¹Max-Planck-Institute for Chemical Physics of Solids, D-01187 Dresden, Germany

²Walther Meissner Institute for Low Temperature Research of the Bavarian Academy of Sciences, D-85748 Garching, Germany

³SPHT, L'Orme des Merisiers, CEA-Saclay, 91190 Gif-sur-Yvette, France

⁴CMT, Department of Physics and Astronomy, Rutgers University, Piscataway, New Jersey 08854-8019, USA

The point at absolute zero where matter becomes unstable to new forms of order is called a quantum critical point (QCP). The quantum fluctuations between order and disorder^{1–5} that develop at this point induce profound transformations in the finite temperature electronic properties of the material. Magnetic fields are ideal for tuning a material as close as possible to a QCP, where the most intense effects of criticality can be studied. A previous study⁶ on the heavy-electron material YbRh₂Si₂ found that near a field-induced QCP electrons move ever more slowly and scatter off one another with ever increasing probability, as indicated by a divergence to infinity of the electron effective mass and scattering cross-section. But these studies could not shed light on whether these properties were an artefact of the applied field^{7,8}, or a more general feature of field-free QCPs. Here we report that, when germanium-doped YbRh₂Si₂ is tuned away from a chemically induced QCP by magnetic fields, there is a universal behaviour in the temperature dependence of the specific heat and resistivity: the characteristic kinetic energy of electrons is directly proportional to the strength of the applied field. We infer that all ballistic motion of electrons vanishes at a QCP, forming a new class of conductor in which individual electrons decay into collective current-carrying motions of the electron fluid.

Recent work⁶ on the heavy electron material YbRh₂Si₂ (ref. 9) has demonstrated that a magnetic field can be used to probe the heavy electron QCP. This material exhibits a small antiferromagnetic (AFM) ordering temperature $T_N = 70$ mK (Fig. 1a) that is driven to zero by a critical magnetic field $B_c = 0.66$ T (if the field is applied parallel to the crystallographic c axis, perpendicular to the easy magnetic plane)⁶. For $B > B_c$, a field-induced Landau–Fermi liquid (LFL) state characterized by $\Delta\rho = AT^2$ (where $\Delta\rho(T) = \rho(T) - \rho_0$ is the temperature-dependent part of the electrical resistivity) is established below some cross-over temperature $T_0(B)$ which grows linearly with field. The A coefficient, being proportional to the quasiparticle–quasiparticle scattering cross-section, was found to diverge as $A(B) \propto 1/(B - B_c)$ for $B \rightarrow B_c$. Comparative studies of the resistivity and the electronic specific heat $C_{el}(T) = \gamma_0(B)T$ in the field ranges 0.5–4 T ($B \perp c$ with $B_c = 0.06$ T) and 2–6 T ($B \parallel c$) revealed a field-independent ratio A/γ_0^2 slightly smaller than the empirical Kadowaki–Woods ratio¹⁰ that holds for LFL systems. This seemed to suggest a divergence of the effective quasiparticle mass with $1/(B - B_c)^{1/2}$ as $B \rightarrow B_c$. Here we report our observation of the divergence of the quasiparticle mass at a QCP, established very close to $B = 0$.

By alloying YbRh₂Si₂ with germanium, using a nominal concentration $x = 0.05$, we have been able to fine-tune the Néel temperature of this material and the critical field far closer to zero, to a point where we can now reliably probe the zero-field transition using field-tuning. The phase diagram for a high-quality YbRh₂(Si_{0.95}Ge_{0.05})₂ single crystal is shown in Fig. 1b. Non-Fermi-liquid (NFL) behaviour dominates over a funnel-shaped region of the T – B phase diagram down to the lowest accessible temperature of

20 mK. The critical field has been suppressed to as low as $B_c = 0.027$ T ($B \perp c$). As in the undoped material, there is a broad cross-over regime between the NFL and field-polarized LFL regimes with a mean cross-over temperature T_0 that is seen to rise linearly with the field B . Very weak AFM order develops in the $x = 0.05$ sample below $T_N = 20$ mK, as evidenced by the extremely weak anomaly in the electronic specific-heat coefficient (Fig. 2a).

Past experience^{7,8} suggested that a finite field QCP has properties which are qualitatively different to a zero-field transition, shedding doubt on the reliability of these measurements as an indicator of the physics of a quantum phase transition at zero field. However, the zero-field properties of YbRh₂(Si_{1–x}Ge_x)₂ above $T \approx 70$ mK for the undoped ($x = 0$) and doped ($x = 0.05$) crystals are essentially identical (Fig. 2a), suggesting that by suppressing the critical field we are still probing the same QCP. In both compounds, the a.c.-susceptibility follows a temperature dependence $\chi^{-1} \propto T^\alpha$ from 0.3 K to 1.5 K, with $\alpha = 0.75$ (ref. 11), and the coefficient of the electronic specific heat, $C_{el}(T)/T$, exhibits⁹ a logarithmic divergence between 0.3 K and 10 K. However, in the low- T paramagnetic regime, that is, $T_N < T \lesssim 0.3$ K, the a.c.-susceptibility follows a Curie–Weiss law (inset of Fig. 2a) with a Weiss temperature

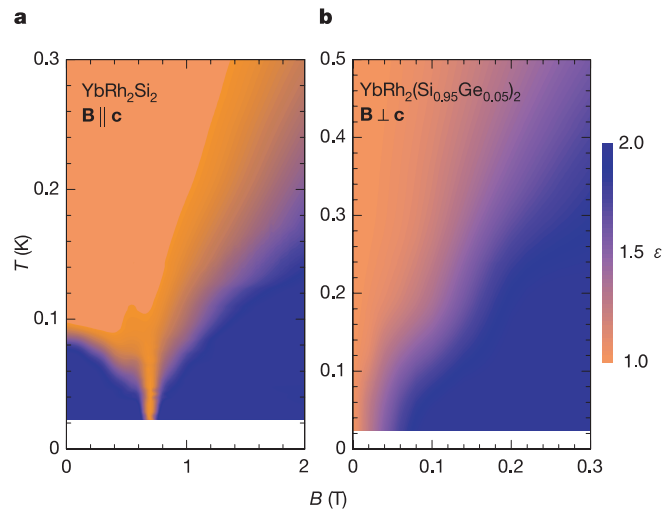


Figure 1 Evolution of ε , the exponent in $\Delta\rho(T) = [\rho(T) - \rho_0] \propto T^\varepsilon$, within the temperature–field phase diagram of YbRh₂(Si_{1–x}Ge_x)₂ single crystals. The non-Fermi-liquid (NFL) behaviour, $\varepsilon = 1$ (yellow), is found to occur at the lowest temperatures right at the QCP, $B = B_c$, and in a largely extended field range at higher temperatures. **a**, $x = 0$, $B_c = 0.66$ T ($B \parallel c$), residual resistivity $\rho_0 = 1 \mu\Omega$ cm; **b**, $x = 0.05$, $B_c = 0.027$ T ($B \perp c$), $\rho_0 = 5 \mu\Omega$ cm. For $B > B_c$, a broad cross-over regime from the NFL state to the field-induced heavy-LFL state (at lower temperature) occurs. The LFL state is characterized by $\Delta\rho(T) \propto T^\varepsilon$, $\varepsilon = 2$ (blue). As shown in **a** the antiferromagnetically ordered phase of pure YbRh₂Si₂ below $T_N = 70$ mK and B_c shows, owing to an extremely small ordered moment, the outward appearance of a heavy LFL state, too. Its phase boundary to the paramagnetic state is manifested by a rapid change in ε from 2 to 1. The low ordering temperature of pure YbRh₂Si₂ increases as external pressure is applied⁹. The extrapolation of $T_N(p) \rightarrow 0$ yields a critical pressure $p_c = -0.3(1)$ GPa, reflecting that a small expansion of the unit cell volume, V , would tune $T_N \rightarrow 0$. This can be achieved by the substitution of Si by the isoelectronic, but larger, Ge (ref. 16). Studies of the electrical resistivity under pressure revealed a $T_N \propto (p + p_c)^n$ variation, with $n = 1.33$ for both compounds. The $T_N(p)$ dependence of the $x = 0$ and $x = 0.05$ crystals can be matched if all $x = 0.05$ data points are shifted by the same amount $\Delta p = -0.17(2)$ GPa to lower pressure¹⁶, yielding $T_N = 20(5)$ mK. Using the bulk modulus $B_0 = 189$ GPa of YbRh₂Si₂ (ref. 17), the small pressure shift of $\Delta p = -0.17(2)$ GPa is equivalent to a volume expansion of $\Delta V = 0.14(3)\text{\AA}^3$. This transforms into an effective Ge content $x_{\text{eff}} = 0.019(6)$, if the value $\Delta V/V = 7.65(78)\%$ for the relative change of the unit-cell volume with Ge concentration in the solid-solution YbRh₂(Si_{1–x}Ge_x)₂ is used, with $V(x = 0) = 158.4(2)\text{\AA}^3$ and $V(x = 1) = 166.07(54)\text{\AA}^3$ (ref. 18) in agreement with microprobe analysis¹¹.

$\Theta \approx -0.3$ K, and a surprisingly large effective moment $\mu_{\text{eff}} \approx 1.4 \mu_B$ per Yb^{3+} , indicating the emergence of coupled, unquenched spins at the QCP. The electronic specific heat coefficient, $C_{\text{el}}(T)/T$, exhibits a pronounced upturn below 0.3 K.

We now discuss the field dependence of the electronic specific heat in $\text{YbRh}_2(\text{Si}_{0.95}\text{Ge}_{0.05})_2$ in more detail. In these measurements, magnetic fields were applied perpendicular to the crystallographic c axis, within the easy magnetic plane (Fig. 2b). At fields above 0.1 T, C_{el}/T is almost temperature-independent, as expected in a LFL¹². A

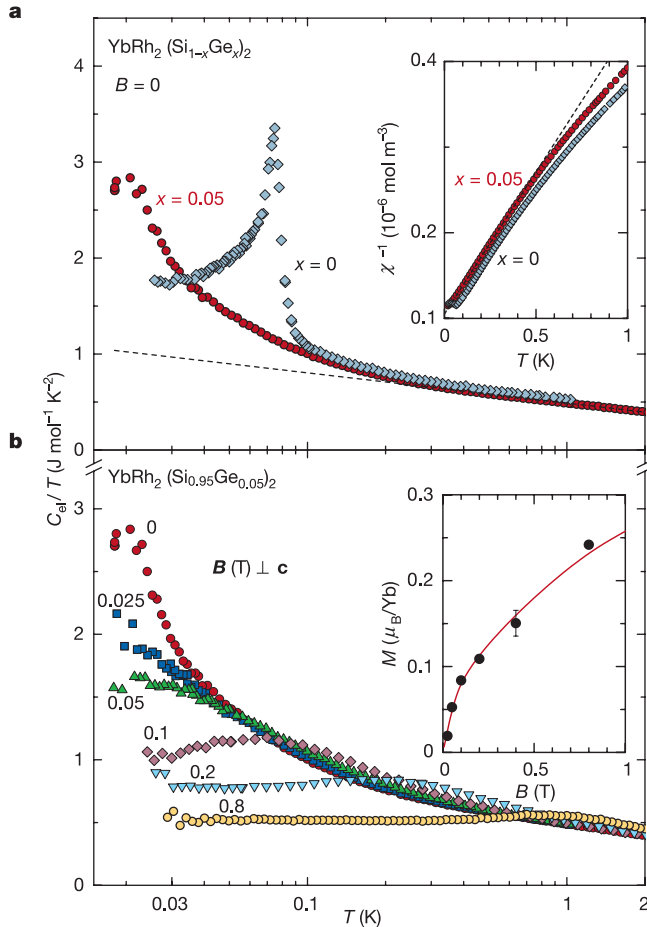


Figure 2 Low-temperature electronic specific heat of $\text{YbRh}_2(\text{Si}_{1-x}\text{Ge}_x)_2$ single crystals as C_{el}/T versus T in semi-logarithmic plots at zero field and at low values of the applied magnetic field B . Insets show low- T , $B = 0$, a.c.-susceptibility as χ^{-1} versus T (a) and magnetization as M versus B (b). C_{el} is obtained by subtracting the nuclear quadrupolar contribution, $C_Q = \alpha_Q/T^2$ (with $\alpha_Q = 5.68 \times 10^{-6} \text{ J K mol}^{-1}$, calculated from recent Mössbauer results¹⁷) (a) and, in addition, the nuclear Zeeman contribution $C_{\text{hf}} = \alpha(B)/T^2$ (b), from the raw data. Here, $\alpha(B)$ has been deduced by plotting CT^2 versus T^3 . The magnetization, M versus magnetic field B (black points in the inset to b), is calculated via $(B_{\text{hf}} - B)/a$, with a hyperfine coupling constant for Yb in this compound and the hyperfine field $B_{\text{hf}} = ((\alpha(B) - \alpha_Q)/\alpha_{\text{dip}})^{1/2}$; α_{dip} represents the strength of the nuclear magnetic dipolar interaction and amounts to $7.58 \times 10^{-8} \text{ J K mol}^{-1} \text{ T}^{-2}$ (ref. 19). With the assumption of $a = 120 \text{ T}/\mu_B$, the data points agree perfectly with the measured magnetization curve at 40 mK (red line in the inset to b). The $B = 0$ results shown in a reveal an upturn in $C_{\text{el}}(T)/T$ for paramagnetic $\text{YbRh}_2(\text{Si}_{1-x}\text{Ge}_x)_2$ ($x = 0$, $T_N = 70 \text{ mK}$; $x = 0.05$, $T_N = 20 \text{ mK}$) below $T = 0.3 \text{ K}$. In the same temperature range the susceptibility $\chi(T)$ shows a Curie–Weiss law, $\chi^{-1} \propto (T - \Theta)$ (inset of a). For both samples very similar values are found for the Weiss temperature, $\Theta \approx -0.3 \text{ K}$, as well as for the large effective moment, $\mu_{\text{eff}} \approx 1.4 \mu_B$ per Yb^{3+} . For $\text{YbRh}_2(\text{Si}_{0.95}\text{Ge}_{0.05})_2$, entropy is shifted from low to higher temperatures when a magnetic field is applied (b). The cross-over temperature between the field-induced LFL state ($C_{\text{el}}(T)/T \approx \text{const.}$) and the NFL state at higher temperature is depicted by the position of the broad hump in $C_{\text{el}}(T)/T$ which shifts upwards linearly with the field, $B \leq 0.8 \text{ T}$ ($B \perp c$).

weak maximum is observed in $C_{\text{el}}(T)/T$ at a characteristic temperature $T_0(B)$ which grows linearly with the field (inset of Fig. 3a), indicating that entropy is transferred from the low-temperature upturn to higher temperatures by the application of a field $B > B_c = 0.027 \text{ T}$. As the field is lowered, the temperature window over which $C_{\text{el}}(T, B)/T = \gamma_0(B)$ is constant shrinks towards zero

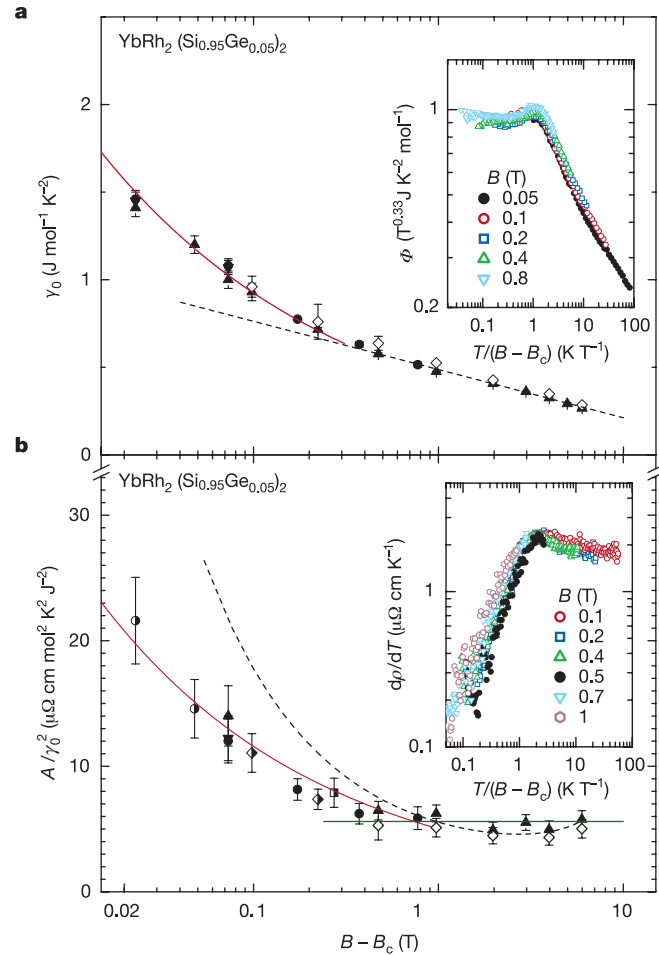


Figure 3 Field dependences of the Sommerfeld coefficient γ_0 , of the electronic specific heat (a) and of the ratio of the A coefficient in the T^2 term of the electrical resistivity and γ_0^2 (b) for $\text{YbRh}_2(\text{Si}_{0.95}\text{Ge}_{0.05})_2$. Note that γ_0 and A are proportional to the effective quasiparticle mass and the effective quasiparticle–quasiparticle scattering cross-section, respectively. The magnetic field was applied perpendicular to the c axis, and the applied field values are corrected, on the abscissae, by the value of the critical field, $B_c = 0.027 \text{ T}$. The γ_0 values in a were obtained from two different samples: three independent measurements on sample 1 are displayed by closed symbols (circles, up and down triangles). The open diamonds show the results of measurements on sample 2. As $B \rightarrow B_c$, γ_0 diverges $\propto (B - B_c)^{-0.33}$ (red line), that is much more strongly than logarithmically (black dashed line). The symbols used in the semi-logarithmic plot $K = A/\gamma_0^2$ versus $(B - B_c)$ of b correspond to values for the electronic specific-heat coefficient shown in a. Half-filled circles (or squares) display data for which the A coefficient of the electrical resistivity was determined by extrapolating (or interpolating) $A(B - B_c)$ with respect to $(B - B_c)$. The half-filled diamond represents a point for which the γ_0 value was obtained by interpolation. The black dashed line indicates $K_{\text{SDW}} \propto [(B - B_c) \ln^2(B - B_c)]^{-1}$ for the 2D SDW scenario¹³. This is at strong variance from the (at $B - B_c < 0.3 \text{ T}$) experimentally observed $K \propto (B - B_c)^{-1/3}$, arising from the stronger than logarithmic increase of γ_0 upon cooling (red line in a). For $(B - B_c) > 0.3 \text{ T}$, K becomes field-independent within the error bars at a constant value of $5.4 \mu\Omega \text{ cm mol}^2 \text{ K}^2 \text{ J}^{-2}$ (green horizontal line in b). A similar high-field behaviour has been reported previously⁶ on pure YbRh_2Si_2 . We note that an almost identical value for K was found. Insets show the scaling behaviour of the low- T electronic specific heat where, according to equation (1), the ordinate is displaying $\Phi(B, T) = (B - B_c)^{0.33} C_{\text{el}}/T$, a, as well as of the temperature derivative of the electrical resistivity, $d\rho/dT$, b, as a function of $T/(B - B_c)$.

and the zero-temperature $\gamma_0(B)$ diverges (Fig. 3a). For example, in a field of 0.05 T a constant value $\gamma_0(B) \approx 1.54(7) \text{ J mol}^{-1} \text{ K}^{-2}$ only develops below 40 mK. These results indicate the formation of a field-induced LFL state at a characteristic scale $T \lesssim T_0(B)$. As the window of LFL behaviour is reduced towards zero, an ever-increasing component of the zero-field upturn in the specific heat coefficient is revealed in the temperature dependence. This confirms that the major part of the upturn in the specific heat coefficient observed in zero field is electronic in character, and must be associated with the intrinsic specific heat at the QCP.

This conclusion is also supported by the electrical resistivity data which reveal a field-dependent cross-over from a T -linear resistivity at elevated temperatures, to quadratic behaviour $\Delta\rho = A(B)T^2$ at sufficiently low temperatures. Most importantly, the data show that at low fields and temperatures, the same scale $T_0(b) \propto b$ (where $b = B - B_c$ is the deviation from the critical field) governs the cross-over from LFL to NFL behaviour in both the thermodynamics and the resistivity. This can be quantitatively demonstrated by noting that the finite field transport and specific-heat data collapse into a single set of scaling relations (see Fig. 3 insets):

$$\frac{C_V}{T} = \frac{1}{b^{1/3}} \Phi\left(\frac{T}{T_0(b)}\right), \quad \frac{d\rho}{dT} = F\left(\frac{T}{T_0(b)}\right) \quad (1)$$

where $\Phi(x) \approx (\max(x, 1))^{-1/3}$ and $F(x) \approx x/\max(x, 1)$. The NFL physics is described by the $x \rightarrow \infty$ ($T \gg T_0(b)$) behaviour of these equations, where $d\rho/dT$ is constant and $C_V/T \propto T^{-1/3}$. By contrast, the field-tuned LFL is described by the $x \rightarrow 0$ limit of these equations. Were there any residual pockets of LFL behaviour that were left unaffected by the QCP, we would expect a residual quadratic component in the resistivity, and the data would not collapse in the observed fashion. We are thus led to believe that the break-up of the LFL involves the entire Fermi surface.

From the second scaling relation in equation (1), we see that the A coefficient of the T^2 term to the resistivity diverges roughly with $1/b$, a result that is consistent with earlier measurements on pure YbRh_2Si_2 carried out further away from the QCP (ref. 6). Over the same range, the specific-heat coefficient $\gamma_0(b)$ grows as $b^{-1/3}$ (Fig. 3a). Notice that the field dependence at absolute zero temperature can be interchanged with the temperature dependence at $B = B_c$, but only in the upturn region. At high magnetic field deviations from the QCP, earlier measurement showed⁶ that the Kadowaki–Woods ratio¹⁰ $K = A/\gamma_0^2$ is approximately constant. Closer to the QCP, where the scaling behaviour is observed, $K = A/\gamma_0^2 \approx b^{-1/3}$ is found to contain a weak field dependence (Fig. 3b).

We now discuss the broader implications of our measurements. The observed divergence of both the A coefficient of the resistivity and the coefficient γ_0 of the T -linear specific heat certainly rule out a three-dimensional spin-density-wave (SDW) scenario, which predicts that both quantities will remain finite at sufficiently low temperature in the approach to a zero-field QCP ($B \rightarrow B_c \rightarrow 0$), but it can be used to gain insight into the underlying scattering mechanisms between the quasiparticles. In a two-dimensional (2D) SDW scenario, the scattering amplitude between two heavy electrons is severely momentum-dependent. When used to compute the transport relaxation rate, the SDW scenario leads to the result $A \propto 1/\kappa^2$, with κ the inverse correlation length¹³. The observed divergence in $A(b)$ would require $\kappa^2 \propto b$. However, the fluctuations of the soft 2D spin fluctuations produce only a weak logarithmic renormalization in the heavy electron density of states, measured by the specific-heat coefficient, $\gamma_0 \propto \ln(1/\kappa)$. Thus the 2D SDW scenario predicts a weak divergence in the T -linear specific heat, but a strongly field-dependent enhancement of the Kadowaki–Woods ratio in the approach to a QCP ($b \rightarrow 0$), given by:

$$\gamma_{\text{SDW}} \propto \ln(1/b) \quad \text{and} \quad K_{\text{SDW}} \propto \frac{1}{b \ln^2(b)}. \quad (2)$$

The strong violation of these predictions in our data, presented in

Fig. 3, rules out 2D spin fluctuations as the driving force behind the thermodynamics and the dominant source of scattering near the heavy electron QCP.

Taking a more general view, scaling behaviour of the transport scattering rate tells us that the only scale entering into the density of states and the scattering amplitude is the single scale $T_0 \propto b$ of the heavy electron fluid. A truly field-independent Kadowaki–Woods ratio would indicate that the quasiparticle scattering amplitude has the form:

$$A^* = T_F^* \mathcal{F}[\{k_{\text{in}}\} \rightarrow \{k_{\text{out}}\}]. \quad (3)$$

The weak field dependence of the Kadowaki–Woods ratio over a wide range of fields implies that the characteristic length scale of the most singular scattering amplitudes renormalizes more slowly in the approach to the QCP than expected in a SDW scenario.

Our data also provide some insight into the thermodynamics in the vicinity of the QCP. By integrating the scaling form (1) for the specific heat over temperature, the entropy $S(T) = \int_0^T dT' (C_V/T')$ in the vicinity of the QCP can be described by the form:

$$S(T, B) = b^{1-\eta} \mathcal{S}\left(\frac{T}{T_0(b)}\right) \quad (4)$$

where $\eta = 1/3$. The appearance of a field-dependent pre-factor in this equation forces the entropy to vanish at the QCP, as required by the third law of thermodynamics. The exponent in the pre-factor also determines the effective Fermi temperature $T_F^*(b) \propto \gamma(b)^{-1} \propto T_0(b)^\eta$. Thus the requirement that the entropy vanishes at the QCP ($\eta < 1$) prevents a direct proportionality between the Fermi-temperature of the heavy LFL and the scale $T_0(b)$ governing the cross-over to NFL behaviour. It follows that the Fermi temperature and cut-off temperature $T_0(b)$ must obey a relationship of the form:

$$T_F^*(b) = T_A \left(\frac{T_0(b)}{T_A}\right)^\eta \quad (5)$$

where T_A is an upper cut-off that we might identify with the single-ion Kondo temperature of the Yb^{3+} ions ($\approx 25 \text{ K}$). Such a power-law renormalization of the characteristic energy scale would be expected in the presence of locally critical fluctuations that extend down from T_A to the infrared cut-off provided, in this case, by the magnetic field¹⁴.

In this respect, our results support the conclusions recently drawn from earlier measurements on the quantum critical material $\text{CeCu}_{6-x}\text{Au}_x$ ($x = 0.1$) (ref. 15), and used in a recently proposed theory for quantum criticality⁴, suggesting that the most critical scattering is neither three-, two- or even one-dimensional, but local—as if the most critical fluctuations in the underlying quantum phase transition are fundamentally ‘zero dimensional’ in character.

One of our most striking observations is that below $T \approx 0.3 \text{ K}$ where $\chi(T)$ follows a Curie–Weiss law, the electronic specific-heat coefficient $C_{\text{el}}(T)/T$ for both samples starts to deviate towards larger values, separating away from the $-\log T$ dependence that is valid⁹ up to 10 K (Fig. 2a). This ‘upturn’ continues in the $x = 0.05$ sample down to approximately 20 mK, if the critical field of 0.027 T ($B \perp c$) is applied (Fig. 2b). We ascribe this intrinsically electronic feature to the critical fluctuations associated with the zero-field quantum phase transition that exists at a slightly larger Ge concentration. The unique temperature dependence of $C_{\text{el}}(T)/T$ for $T < 0.3 \text{ K}$ is disparate from the linear temperature dependence of the electrical resistivity which holds all the way from $\approx 10 \text{ K}$ to $T \approx 10 \text{ mK}$. Since the former (thermodynamic) quantity probes the dominating local $4f$ (‘spin’) part of the composite quasiparticles, while the latter (transport) quantity is sensitive to the itinerant conduction-electron (‘charge’) part, one may view the observed disparity as a direct manifestation of the break-up of the composite fermions in the approach to the QCP. $\square$

Received 24 January; accepted 28 May 2003; doi:10.1038/nature01774.

- Hertz, J. A. Quantum critical phenomena. *Phys. Rev. B* **14**, 1165–1184 (1976).
- Millis, A. J. Effect of a nonzero temperature on quantum critical points in itinerant fermion systems. *Phys. Rev. B* **48**, 7183–7196 (1993).
- Continentino, M. A. Quantum scaling in many-body systems. *Phys. Rep.* **239**, 179–213 (1994).
- Si, Q., Rabello, S., Ingersent, K. & Smith, J. L. Locally critical quantum phase transitions in strongly correlated metals. *Nature* **413**, 804–808 (2001).
- Coleman, P. & Pépin, C. What is the fate of the heavy electron at a quantum critical point? *Physica B* **312**, 383–389 (2002).
- Gegenwart, P. *et al.* Magnetic-field induced quantum critical point in YbRh_2Si_2 . *Phys. Rev. Lett.* **89**, 056402 (2002).
- Heuser, K. *et al.* Inducement of non-Fermi-liquid behavior with a magnetic field. *Phys. Rev. B* **57**, R4198–R4201 (1998).
- Stockert, O. *et al.* Pressure versus magnetic-field tuning of a magnetic quantum phase transition. *Physica B* **312–313**, 458–460 (2002).
- Trovarelli, O. *et al.* YbRh_2Si_2 : Pronounced non-Fermi-liquid effects above a low-lying magnetic phase transition. *Phys. Rev. Lett.* **85**, 626–629 (2000).
- Kadowaki, K. & Woods, S. B. Universal relationship of the resistivity and specific heat in heavy-fermion compounds. *Solid State Commun.* **58**, 507–509 (1986).
- Gegenwart, P. *et al.* Divergence of the heavy quasiparticle mass at the antiferromagnetic quantum critical point in YbRh_2Si_2 . *Acta Phys. Pol. B* **34**, 323–334 (2003).
- Landau, L. D. The theory of a Fermi liquid. *Sov. Phys. JETP* **3**, 920–925 (1957).
- Paul, I. & Kotliar, G. Thermoelectric behavior near the magnetic quantum critical point. *Phys. Rev. B* **64**, 184414 (2001).
- Giamarchi, T., Varma, C. M., Ruckenstein, A. E. & Nozières, P. Singular low energy properties of an impurity model with finite range interactions. *Phys. Rev. Lett.* **70**, 3967–3970 (1993).
- Schröder, A. *et al.* Onset of antiferromagnetism in heavy-fermion metals. *Nature* **407**, 351–355 (2000).
- Mederle, S. *et al.* Unconventional metallic state in $\text{YbRh}_2(\text{Si}_{1-x}\text{Ge}_x)_2$ —a high pressure study. *J. Phys. Condens. Matter* **14**, 10731–10736 (2002).
- Plessel, J. *et al.* Unusual behavior of the low-moment magnetic ground-state of YbRh_2Si_2 under high pressure. *Phys. Rev. B* **67**, 180303 (2003).
- Francois, M., Venturini, G., Marchéché, J. F., Malaman, B. & Roques, B. De nouvelles séries de germaniures, isotopes de $\text{U}_4\text{Re}_2\text{Si}_6$, ThCr_2Si_2 et CaBe_2Ge_2 , dans les systèmes ternaires R-T-Ge où R est un élément des terres rares et T = Ru, Os, Rh, Ir: supraconductivité de LaIr_2Ge_2 . *J. Less Common Metals* **113**, 231–237 (1985).
- Carter, G. C., *et al.* in *Metallic shifts in NMR*. *Progress in Materials Science* Vol. 20, Part I, Ch. 9 (eds Chalmers, B., Christian, J. W. & Massalski, T. B.) 123–124 (Oxford, Pergamon, 1977).

Acknowledgements We acknowledge discussions with J. Ferstl, C. Langhammer, S. Mederle, N. Oeschler, I. Zerec, G. Sparr, O. Stockert, M. Abd-Elmeguid, J. Hopkinson, A. I. Larkin and I. Paul. Work at Dresden is partially supported by the Fonds der Chemischen Industrie and by the FERLIN project of the European Science Foundation. P. C. is supported by the National Science Foundation. Y. T. is a Young Scientist Research Fellow supported by the Japan Society for the Promotion of Science.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.C. (coleman@physics.rutgers.edu).

Superconductivity phase diagram of $\text{Na}_x\text{CoO}_2 \cdot 1.3\text{H}_2\text{O}$

R. E. Schaak¹, T. Klimczuk^{1,2}, M. L. Foo¹ & R. J. Cava^{1,3}

¹Department of Chemistry, Princeton University, Princeton, New Jersey 08544, USA

²Faculty of Applied Physics and Mathematics, Gdansk University of Technology, Narutowicza 11/12, 80-952 Gdansk, Poland

³Princeton Materials Institute, Princeton University, Princeton, New Jersey 08540, USA

The microscopic origin of superconductivity in the high-transition-temperature (high- T_c) copper oxides remains the subject of active inquiry; several of their electronic characteristics are well established as universal to all the known materials, forming the experimental foundation that all theories must address. The most fundamental of those characteristics, for both the copper oxides and other superconductors, is the dependence of the superconducting T_c on the degree of electronic band filling. The recent report of superconductivity¹ near 4 K in the layered sodium cobalt oxyhydrate, $\text{Na}_{0.35}\text{CoO}_2 \cdot 1.3\text{H}_2\text{O}$, is of interest

owing to both its triangular cobalt–oxygen lattice and its generally analogous chemical and structural relationships to the copper oxide superconductors. Here we show that the superconducting T_c of this compound displays the same kind of behaviour on chemical doping that is observed in the high- T_c copper oxides. Specifically, the optimal superconducting T_c occurs in a narrow range of sodium concentrations (and therefore electron concentrations) and decreases for both underdoped and overdoped materials, as observed in the phase diagram of the copper oxide superconductors. The analogy is not perfect, however, suggesting that $\text{Na}_x\text{CoO}_2 \cdot 1.3\text{H}_2\text{O}$, with its triangular lattice geometry and special magnetic characteristics, may provide insights into systems where coupled charge and spin dynamics play an essential role in leading to superconductivity.

Like the high- T_c superconductors, the $\text{Na}_x\text{CoO}_2 \cdot 1.3\text{H}_2\text{O}$ crystal structure¹ consists of electronically active planes (in this case, edge-sharing CoO_6 octahedra) separated by layers (in this case, $\text{Na}_x \cdot 1.3\text{H}_2\text{O}$) that act as spacers, to yield electronic two-dimensionality, and also act as charge reservoirs (see below). We have found that varying the Na content in $\text{Na}_x\text{CoO}_2 \cdot 1.3\text{H}_2\text{O}$ results in the same type of out-of-plane chemical doping control of in-plane electronic

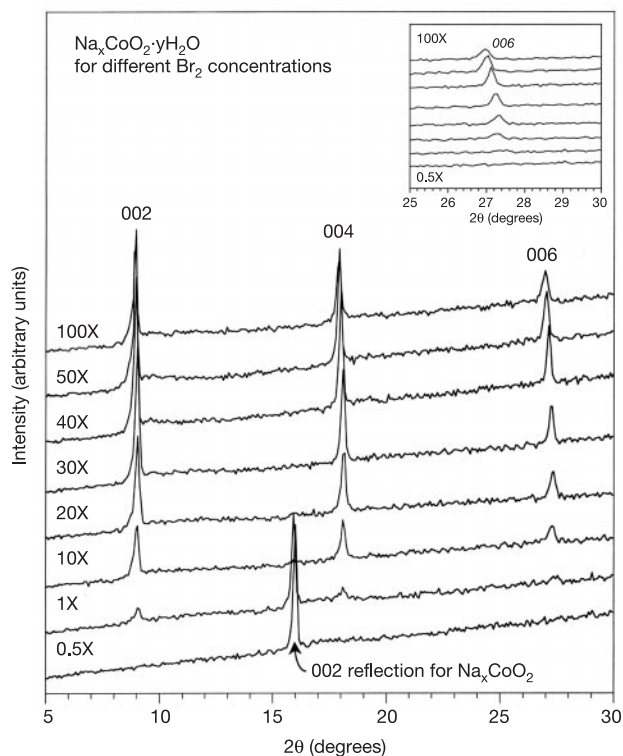


Figure 1 Powder X-ray diffraction patterns (Cu $K\alpha$ radiation) for $\text{Na}_x\text{CoO}_2 \cdot y\text{H}_2\text{O}$ samples prepared using different concentrations of the bromine de-intercalant. Inset, an enlargement of the 006 reflections for each sample, highlighting the shift in the layer spacing as a function of sodium content. The $\text{Na}_x\text{CoO}_2 \cdot y\text{H}_2\text{O}$ samples were prepared by chemically de-intercalating sodium from $\text{Na}_{0.7}\text{CoO}_2$ using bromine as an oxidizing agent. $\text{Na}_{0.7}\text{CoO}_2$ (0.5 g) was stirred in 20 ml of a Br_2 solution in acetonitrile at room temperature for five days. Bromine concentrations representing substoichiometric (0.5X), stoichiometric (1X), and molar excess (10–100X) relative to sodium content were employed. ('1X' indicates that the amount of Br_2 used is exactly the amount that would theoretically be needed to remove all of the sodium from $\text{Na}_{0.7}\text{CoO}_2$). The product was washed several times with acetonitrile and then water, and then dried briefly under ambient conditions. The sodium content of the phases was determined by the inductively coupled plasma atomic emission spectroscopy (ICP-AES) method. Very high Na diffusion coefficients facilitate homogenization of the Na contents of the samples at ambient temperature.

charge that is found for the copper oxide superconductors. This is achieved by changing the Br concentration used in the de-intercalation of the host material. (See Fig. 1 legend for the synthesis procedure). Powder X-ray diffraction (XRD) patterns for the synthesized samples are shown in Fig. 1. The bromine-treated samples made with substoichiometric (0.5X; see Fig. 1 legend for definition of this nomenclature) and stoichiometric (1X) bromine solutions consist primarily of a partially de-intercalated, anhydrous, non-superconducting Na_xCoO_2 phase ($c \approx 11.2 \text{ \AA}$). A small amount of the hydrated superconducting phase $\text{Na}_x\text{CoO}_2 \cdot 1.3\text{H}_2\text{O}$ ($c \approx 19.6 \text{ \AA}$) is detectable by XRD for the 1X sample. Single phase, superconducting, fully hydrated $\text{Na}_x\text{CoO}_2 \cdot 1.3\text{H}_2\text{O}$ occurs for higher Br concentrations, with a small amount of Na_xCoO_2 in the 10X sample. Chemical analysis indicates that the sodium content of the resulting materials generally varies systematically in the samples prepared in different Br concentrations, over a range of $x = 0.26$ to 0.45 (Table 1).

Thermogravimetric analysis of all the samples on very slow heating in oxygen (Fig. 2 inset) showed that their behaviour was identical to that reported previously for $\text{Na}_{0.3}\text{CoO}_2 \cdot y\text{H}_2\text{O}$ (ref. 2). The interlayer 'crystal water' content remains essentially constant, at approximately 1.3 per formula unit, despite differences in sodium content in the single phase samples. The intergrain water is variable from sample to sample, as is to be expected. Figure 1 shows a noticeable shift in the positions of the 001 reflections for the fully hydrated $\text{Na}_x\text{CoO}_2 \cdot 1.3\text{H}_2\text{O}$ phases, yielding a systematic variation in the c axes of the unit cells (Table 1) from 19.43 Å for the $x = 0.45$ sample to 19.77 Å for the $x = 0.26$ sample. This increase in layer separation with decreasing sodium content for the hydrated superconducting phase is similar to that observed in the dehydrated Na_xCoO_2 phase^{3,4}. The a axis, reflecting the in-plane CoO_6 dimensions, is independent of Na content within the precision of our measurements.

Zero field cooled d.c. magnetization data measured in a field of 5 Oe for selected samples are shown in the main panel of Fig. 2. The magnetizations at 1.8 K represent approximately 100% of the theoretical value expected for perfect diamagnetism. Such strong diamagnetic signals provide evidence for bulk superconductivity. Na concentration inhomogeneities in the samples are probably the primary source of rounding of the superconducting transitions. An important point revealed by the data in Fig. 2 is that T_c for each sample is clearly different, indicating that differences in sodium content significantly affect the superconductivity of $\text{Na}_x\text{CoO}_2 \cdot 1.3\text{H}_2\text{O}$.

In order to characterize fully the dependence of T_c on sodium content, we used a.c. susceptibility, which is more sensitive to weakly superconducting samples. Figure 3 shows the a.c. susceptibility for all the $\text{Na}_x\text{CoO}_2 \cdot y\text{H}_2\text{O}$ samples. For the multiple phase $x = 0.45$ and $x = 0.40$ samples, the T_c values are approximately 2.0 K (Table 1), and the diamagnetic a.c. signals are very small,

Table 1 Characterization of $\text{Na}_x\text{CoO}_2 \cdot y\text{H}_2\text{O}$

Sodium content* (x in $\text{Na}_x\text{CoO}_2 \cdot y\text{H}_2\text{O}$)	Bromine concentration	a axis of $\text{Na}_x\text{CoO}_2 \cdot y\text{H}_2\text{O}$ † (Å)	c axis of $\text{Na}_x\text{CoO}_2 \cdot y\text{H}_2\text{O}$ † (Å)	T_c ‡ (K)
0.45	0.5 x	N/A	N/A	2.0
0.40	1 x	2.823(3)	19.43(2)	2.0
0.32	10 x	2.825(2)	19.52(2)	2.1
0.33	20 x	2.823(2)	19.58(1)	2.2
0.32	30 x	2.822(2)	19.58(1)	3.0
0.30	40 x	2.823(2)	19.69(2)	4.3
0.29	100 x	2.819(3)	19.77(2)	4.0
0.26	50 x	2.821(2)	19.77(2)	2.4

$\text{Na}_x\text{CoO}_2 \cdot y\text{H}_2\text{O}$ was prepared by bromine de-intercalation and hydration of $\text{Na}_{0.7}\text{CoO}_2$.

* Sodium content determined by ICP-AES. The estimated error of analysis is ± 0.02 per formula unit.

† Determined by least squares refinement of powder XRD data, from 6–10 reflections between 2θ values of 5° and 60° .

‡ T_c s determined from the a.c. susceptibility data, from the extrapolation of the steepest slope of the M versus T curves in Fig. 3 to $M = 0$.

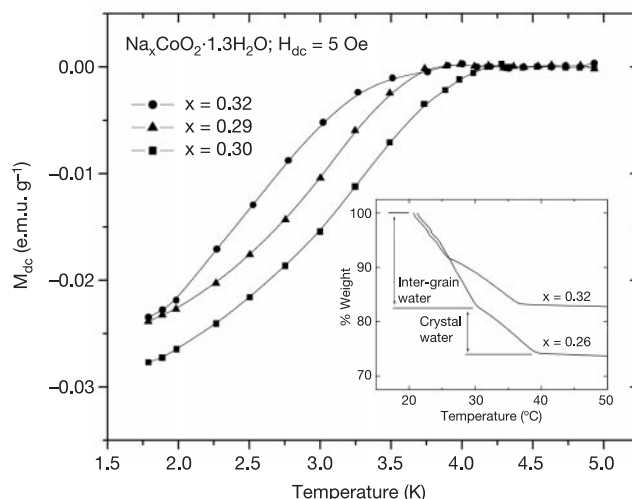


Figure 2 Zero field cooled d.c. magnetization for superconducting samples of $\text{Na}_x\text{CoO}_2 \cdot 1.3\text{H}_2\text{O}$. Data were obtained for $x = 0.29, 0.30$ and 0.32 , using a Quantum Design PPMS magnetometer, $H_{dc} = 5 \text{ Oe}$. Inset, loss in weight of single phase $\text{Na}_x\text{CoO}_2 \cdot 1.3\text{H}_2\text{O}$ ($x = 0.26$ and 0.32) samples heated extremely slowly in O_2 (0.25 degrees per minute) illustrating the method by which we distinguish the amount of crystal water (the higher-temperature weight loss) from the intergrain water (the lower-temperature weight loss). The change in weight that occurs on loss of crystal water is seen to be essentially the same in both low-Na and high-Na content materials.

consistent with their phase analysis by XRD, which shows primarily non-superconducting, anhydrous Na_xCoO_2 . In the $x = 0.40$ sample, the fully hydrated $\text{Na}_x\text{CoO}_2 \cdot 1.3\text{H}_2\text{O}$ phase accounts for approximately 15% of the sample, allowing us to estimate the maximum sodium content of the $\text{Na}_x\text{CoO}_2 \cdot 1.3\text{H}_2\text{O}$ phase to be approximately $x = 0.35$. The data suggest that at its highest possible Na content, the $\text{Na}_x\text{CoO}_2 \cdot 1.3\text{H}_2\text{O}$ phase has a T_c near 2 K.

All other samples are single phase sodium cobalt oxyhydrate with the crystal structure of the superconductor. The $x = 0.32$ and $x = 0.33$ samples yield slightly higher T_c values (between 2.1 and 2.2 K) and signals that are one to two orders of magnitude higher than the multiple phase $x = 0.40$ sample. Single phase samples with sodium contents of $x = 0.32, 0.30$ and 0.29 display superconducting

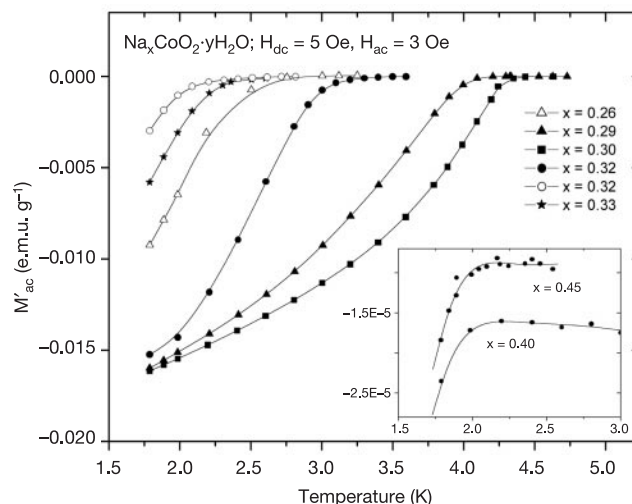


Figure 3 Zero-field cooled a.c. magnetization for all superconducting $\text{Na}_x\text{CoO}_2 \cdot y\text{H}_2\text{O}$ samples. ($H_{dc} = 3 \text{ Oe}$, $H_{ac} = 5 \text{ Oe}$, $f = 10 \text{ kHz}$) Magnetization data for the weakly superconducting samples $x = 0.45$ and 0.40 are shown in the inset.

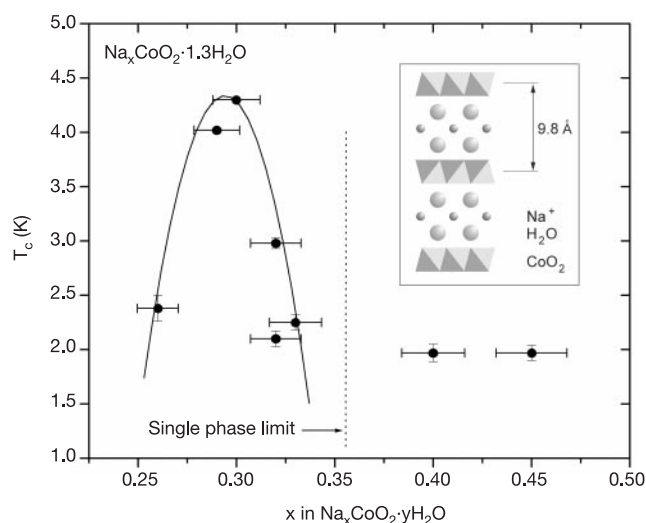


Figure 4 The superconducting phase diagram for $\text{Na}_x\text{CoO}_2 \cdot 1.3\text{H}_2\text{O}$. Main panel, T_c as a function of x as determined from the a.c. susceptibility measurements in Fig. 3. Inset, schematic representation of the layered crystal structure of $\text{Na}_x\text{CoO}_2 \cdot 1.3\text{H}_2\text{O}$. Triangular layers of CoO_6 edge-shared octahedra are shown in a polyhedral representation.

T_c values of 3.0 K, 4.3 K and 4.0 K, respectively (Fig. 3). Significantly, the single phase sample with $x = 0.26$ has a T_c of only 2.4 K.

Figure 4 shows, in the main panel, the superconducting phase diagram of $\text{Na}_x\text{CoO}_2 \cdot y\text{H}_2\text{O}$. The variation of T_c with x indicates that there is an optimal sodium composition for the occurrence of superconductivity, $x = 0.30$. T_c decreases at both lower and higher Na contents. As the sodium content increases between $x = 0.26$ and $x = 0.35$, the formal oxidation state of the Co decreases from 3.74+ to 3.65+, and the optimum for superconductivity is 3.70+. This variation of T_c with the degree of electronic doping of the CoO_2 planes, is analogous with the behaviour observed in the copper oxide superconductors. We note that the synthesis method we have employed, and other ambient temperature synthesis methods that might be used in this system, are likely to result in a distribution of sodium contents for each sample. If ideally uniform sodium content samples can be prepared, we expect that the superconducting 'dome' shown in Fig. 4 may become more narrow in sodium content.

Preliminary correlation of the chemical doping due to the Na content and the true electronic doping state of the CoO_2 planes can be accomplished by electron counting in the context of electronic pictures already being developed for both the dehydrated Na_xCoO_2 and $\text{Na}_x\text{CoO}_2 \cdot 1.3\text{H}_2\text{O}$ phases (see, for example, refs 4–9). For $x = 0$, the formal Co oxidation state is Co 4+, with a t_{2g}^5 electron configuration in the low spin state. For $x = 1$, Co is formally 3+, with an electron configuration of t_{2g}^6 in the low spin state. Electronic structure calculations for $\text{Na}_{0.5}\text{CoO}_2$ (ref. 6) indicate that the t_{2g} band would be completely filled at $x = 1$, and that an 'ordinary' semiconductor is expected. However, it has been pointed out that the trigonal distortion of the CoO_6 octahedra in these structures will probably lead to splitting of the t_{2g} band^{5,10}. The proposed t_{2g} band splitting would then, for $x = 0$, result in a fully filled four-electron band and a half-filled two-electron band, leading to the expectation that this compound will be a Mott–Hubbard insulator. $\text{Na}_x\text{CoO}_2 \cdot 1.3\text{H}_2\text{O}$ for $x = 0$ would then bear a striking similarity to La_2CuO_4 , the ground state for the copper oxide superconductors, where electrons at the Fermi energy also reside in a half-filled two-electron band. It has not yet been determined whether this half-filled state in the layered triangular lattice cobalt oxides at $x = 0$ gives rise to a Mott–Hubbard insulator as it does in the copper oxides. In this scenario, then, each added Na above $x = 0$ in $\text{Na}_x\text{CoO}_2 \cdot 1.3\text{H}_2\text{O}$ results in the addition of one electron per

cobalt to the half-filled band, and the optimal chemical doping level for superconductivity, $x = 0.3$, represents the addition of 0.3 electrons to the half-filled band per formula unit. Values of x less than 0.3 would then represent underdoped materials, and values of x greater than 0.3 would represent overdoped materials. More detailed characterization of the electronic state of the triangular cobalt oxides will be needed to determine just how closely the electronic analogies to the copper oxides hold.

This work reveals several experimental findings that are critical for understanding the superconductivity of $\text{Na}_x\text{CoO}_2 \cdot 1.3\text{H}_2\text{O}$. The discovery of a maximum in T_c as a function of the Na content establishes the optimal chemical doping level for superconductivity. A fundamental similarity between the layered copper oxide and layered cobalt oxide superconductors is seen in the decrease in T_c for both the underdoped and overdoped materials. The optimal doping level for superconductivity is clearly higher with respect to the Mott–Hubbard-like half-filled two-electron band in $\text{Na}_x\text{CoO}_2 \cdot 1.3\text{H}_2\text{O}$ (+0.3 electrons) than in the copper oxides (+0.15 electrons or holes). This may either reflect a fundamental difference between the two types of superconductors, or the influence of as-yet unknown structure–electronic state correlations in the cobalt oxides. Observations of unusual electrical transport properties in the host material $\text{Na}_{0.7}\text{CoO}_2$ itself suggest that coupled spin and charge dynamics may be implicated in the superconductivity, although there are significant differences from copper oxide systems^{10,11}. Observations that the lower hydrates with closer CoO_2 – CoO_2 interplanar distances are not superconducting above 2 K (ref. 2), and that T_c decreases under pressure¹², indicate that the two-dimensional character of the structure is important. Though the intrinsically complex materials chemistry of the $\text{Na}_x\text{CoO}_2 \cdot 1.3\text{H}_2\text{O}$ superconductor makes it difficult to characterize, we believe that potentially fruitful comparisons to the copper oxides, and the fact that this compound may represent the literal embodiment of Anderson's original proposal for the resonating valence bond state¹³, make it worthy of further study. □

Received 6 June; accepted 2 July 2003; doi:10.1038/nature01877.

1. Takada, K. *et al.* Superconductivity in two-dimensional CoO_2 layers. *Nature* **422**, 53–55 (2003).
2. Foo, M. L. *et al.* Chemical instability of the cobalt oxyhydrate superconductor under ambient conditions. *Solid State Commun* **127**, 33–37 (2003).
3. Fouassier, C., Matjeka, G., Reau, J.-M. & Hagenmüller, P. Sur de nouveaux bronzes oxygènes de formule Na_xCoO_2 ($x \leq 1$). Le système cobalt-oxygène-sodium. *J. Solid State Chem.* **6**, 532–537 (1973).
4. Baskaran, G. An electronic model for CoO_2 layer based systems: Chiral RVB metal and superconductivity. Preprint at (<http://xxx.lanl.gov/cond-mat/0303649>) (2003).
5. Kumar, B. & Shastri, B. S. Superconductivity in CoO_2 layers and the resonating valence bond mean field theory of the triangular lattice t-J model. Preprint at (<http://xxx.lanl.gov/cond-mat/0304210>) (2003).
6. Singh, D. J. Electronic structure of NaCoO_2 . *Phys. Rev. B* **61**, 13397–13402 (2000).
7. Honeramp, C. Instabilities of interacting electrons on the triangular lattice. Preprint at (<http://xxx.lanl.gov/cond-mat/0304460>) (2003).
8. Tanaka, A. & Hu, X. Possible spin triplet superconductivity in $\text{Na}_x\text{CoO}_2 \cdot y\text{H}_2\text{O}$. Preprint at (<http://xxx.lanl.gov/cond-mat/0304409>) (2003).
9. Wang, Q.-H., Lee, D.-H. & Lee, P. A. Doped t-J model on a triangular lattice: Possible application to $\text{Na}_x\text{CoO}_2 \cdot y\text{H}_2\text{O}$ and Na_xTiO_2 . Preprint at (<http://xxx.lanl.gov/cond-mat/0304377>) (2003).
10. Terasaki, I., Sasago, Y. & Uchinokura, K. Large thermoelectric power in NaCoO_2 single crystals. *Phys. Rev. B* **56**, R12685–R12687 (1997).
11. Wang, Y., Rogado, N. S., Cava, R. J. & Ong, N. P. Spin entropy as the likely source of enhanced thermopower in NaCoO_2 . *Nature* **423**, 425–428 (2003).
12. Lorenz, B., Cmaidalka, J., Meng, R. L. & Chu, C. W. Effect of hydrostatic pressure on the superconductivity in $\text{Na}_x\text{CoO}_2 \cdot y\text{H}_2\text{O}$. Preprint at (<http://xxx.lanl.gov/cond-mat/0304537>) (2003).
13. Anderson, P. W. Resonating valence bonds: A new type of insulator. *Mater. Res. Bull.* **8**, 153–160 (1973).

Acknowledgements This work was supported by the US National Science Foundation, Division of Materials Research, and the US Department of Energy, Division of Basic Energy Sciences. T.K. thanks the Foundation for Polish Science for support.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.J.C. (rcava@princeton.edu).

A recyclable catalyst that precipitates at the end of the reaction

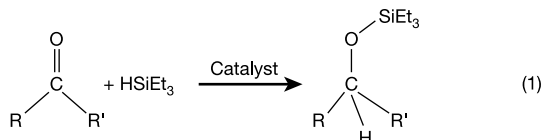
Vladimir K. Dioumaev & R. Morris Bullock

Chemistry Department, Brookhaven National Laboratory, Upton, New York 11973-5000, USA

Homogeneous catalysts—which exist in the same (usually liquid) phase as reactants and products—are usually more selective than heterogeneous catalysts and far less affected by limitations due to slow transport of reactants and products, but their separation from reaction products can be costly and inefficient. This has stimulated the development of strategies that facilitate the recycling of homogeneous catalysts^{1–4}. Some of these methods exploit the preference of a catalyst for one of two solvents with thermoregulated miscibility^{5,6}; others exploit a dramatic decrease in catalyst solubility as one reagent is consumed^{7,8} or temperature changed after completion of the reaction^{9–14}. Here we describe a tungsten catalyst for the solvent-free hydrosilylation of ketones that retains its activity until essentially all of the liquid substrate is converted to liquid products, which we can then simply decant to separate the catalyst that precipitates from the products of the reaction. We attribute the ability of the catalyst to retain its solubility and hence activity until completion of the reaction to the transient formation of liquid clathrate^{15,16} that contains a few molecules of the substrate per molecule of the otherwise solid catalyst. Insights into the fundamental processes controlling the formation of this liquid clathrate might help to tailor other catalysts and substrates, so as to develop efficient and solvent-free schemes for reactions of practical interest.

The principles of 'green' chemistry¹⁷ and 'green' engineering¹⁸ dictate that avoiding the use of solvents¹⁹ is an important way to prevent generation of waste. Furthermore, a solvent-free transformation from pure reagents to pure products potentially yields a dramatic change in the properties of the medium and offers a tantalizing opportunity for attaining catalyst self-precipitation. Precipitation, in turn, helps to avoid using solvents in the subsequent separation stages, further preventing waste generation. Whereas the concept is trivial, striking the right balance between maintaining catalyst solubility throughout the reaction and precipitation at the end is a very difficult problem. An ideal catalyst should stay at least somewhat soluble until the last molecule of the substrate is consumed. Rare instances of such retention of solubility are known among compounds with a low aptitude for crystal lattice formation. Such compounds can furnish a liquid phase—a liquid clathrate^{15,16}—with just a few equivalents of the solvent per equivalent of the otherwise solid component²⁰. This behaviour is observed among ionic complexes with weakly coordinating counterions, where the charges are delocalized over large molecular fragments, and the crystal packing forces are weakened^{21–23}.

To explore the concept of self-separation, two cationic complexes with weakly coordinating B(C₆F₅)₄[–] anions have been probed as catalysts for hydrosilylation of carbonyl compounds: [CpM(CO)₂(IMes)]⁺[B(C₆F₅)₄][–] (where Cp = cyclopentadienyl, IMes = 1,3-bis(2,4,6-trimethylphenyl)-imidazol-2-ylidene; in **1Mo** M = Mo, in **1W** M = W). Hydrosilylation is a suitable model reaction as it starts with a polar liquid substrate (ketone or ester) and ends with a non-polar liquid product (alkoxysilane, equation (1) below).



Hydrosilylation is also a reaction of considerable practical interest as it is widely used for both large- and small-scale syntheses^{24,25}. Not only are alkoxysilanes precursors to silicon-containing polymers and ceramic materials, but they are also valuable in organic synthesis. Thus, for the conversion of carbonyl compounds to alcohols, hydrosilylation is often used as a convenient alternative to hydrogenation, particularly in asymmetric synthesis.

The choice of catalysts is based on our prior experience with **1Mo** and **1W** in the hydrogenation of ketones²⁶, a reaction which is mechanistically related to hydrosilylation and is often catalysed by similar complexes. We have previously found **1Mo** and **1W** to be soluble in ketones, insoluble in non-polar hydrocarbon solvents, and prone to the formation of oily precipitates—liquid clathrates—instead of crystalline products. We have also found that the *N*-heterocyclic carbene ligand, IMes, stabilizes the electronically unsaturated 16e[–] **1Mo** and **1W** complexes by formation of a weak bond between the metal centre and one of the C=C double bonds of a mesityl group²⁶. Although stabilization of catalytic intermediates inhibits catalysis, moderate stabilization is required for recycling, as the catalyst and its resting states should be able to withstand recycling conditions and should show reasonable tolerance to common impurities in the system.

Complexes **1W** and **1Mo** indeed catalyse hydrosilylation of carbonyl compounds under mild conditions, and the tungsten

Table 1 Hydrosilylation of carbonyl compounds by W catalysts

Substrate	Products	Initial TOF (h ^{–1})	Total TON	yield (%)	Time (h)
		370	446	89	1
		20	36	7	
		~5	15	3	
		>2,000	447	93	0.25*
		>100	24	5	
		150	489	98	19
		110	446	89	23
		<5	11	2	
		170	468	94	26†
		<1	30	6	
		30	386	77	168
		<1	70	14	

Reactions were conducted at 23 °C in pure liquid substrates without solvent: ketone / HSiEt₃ / **1W** = 100 / 120 / 0.2. Turnover number (TON) is the number of moles of a carbonyl substrate consumed to yield a given product per the number of moles of catalyst. TOF is the average initial turnover frequency measured within the first 15–20 min of the reaction (TOF = TON / time).

* HSiMe₂Ph was used instead of HSiEt₃.

† The ratios were: ester / HSiEt₃ / **1W** = 100 / 220 / 0.2.

complex **1W** proves to be much more active than the molybdenum complex **1Mo**. The reactions exhibit good rates, high conversions, and an excellent selectivity for hydrosilylation of C=O versus C=C double bonds (Table 1 and Supplementary Information). Hydrosilylation of aromatic substrates yields a brown oily precipitate towards the end of the reaction, but some of the catalyst remains soluble. Similarly, hydrosilylation with an aromatic silane, Me₂PhSiH, also results in a partially soluble catalyst and a small amount of a brown oily precipitate. Aliphatic substrates, on the other hand, yield colourless solutions at the end of the reaction with no detectable soluble metal-containing species in the proton nuclear magnetic resonance (¹H NMR) spectra. Conversion of the last traces of the carbonyl substrate can be monitored visually as the precipitate transforms from a purple oil into a pale yellow solid (Fig. 1). Fortunately, the precipitate is somewhat sticky, and can be readily recovered by decanting the liquid products without any special precautions to retain the catalyst, and no solvent is needed for the reaction work-up.

The actual resting state of the tungsten catalyst that is recycled has been found to be a mixture of [CpW(CO)₂(IMes)(SiEt₃H)]⁺[B(C₆F₅)₄][−] (**2W**) and [CpW(CO)₂(IMes)(H)₂]⁺[B(C₆F₅)₄][−] (**3W**). The assignment is confirmed by the independent syntheses of both compounds from **1W** and HSiEt₃ or dihydrogen (see Supplementary Information for complete characterization). The solubility of **2W** and **3W** in the products of hydrosilylation of Et₂C=O is below the detection limits of ¹H NMR spectroscopy. The residual solubility of all species with a B(C₆F₅)₄[−] counter-ion has been measured by fluorine nuclear magnetic resonance (¹⁹F NMR) spectroscopy to be ~4 × 10^{−4} mol l^{−1}, which corresponds to ~5% of the loaded 0.2 mol% catalyst. In other words, more than 95% of the loaded catalyst has precipitated as the resting state, and is amenable for recycling. The recovered catalyst exhibits up to twice the activity after the first recycle, and retains good activity for all five cycles performed (Table 2). Thus, either **2W**, or **3W**, or both, are better catalyst precursors than **1W**.

The ketone complex [CpW(CO)₂(IMes)(Et₂C=O)]⁺[B(C₆F₅)₄][−] (**4W**) is also a resting state present during hydrosilylation of Et₂C=O, giving a purple colour to the reaction mixture (λ_{max}(toluene) = 498 nm, ε = 1 × 10³ l mol^{−1} cm^{−1}). Identified by multinuclear NMR and infrared (IR) spectroscopy, the assignment of **4W** has been confirmed by an independent synthesis from **1W** and Et₂C=O. Complex **4W** is most abundant at the beginning of the hydrosilylation and is gradually replaced by **2W** and **3W**. The formation of the dihydride complex **3W** is due to traces of H₂ produced from HSiEt₃ and residual water. The equilibrium for the

Table 2 Recycling of catalyst for the hydrosilylation of Et₂C=O

Cycle no.	1	2	3	4	5
Time of measurement (min)	15	10	10	10	10
TOF (h ^{−1})	370	780	870	760	620

Reaction conditions and TOF are the same as those in Table 1.

formation of **3W** is very favourable; K_{eq} = [**3W**][Et₂C=O]/[**4W**] × [H₂] ≈ 1 × 10³ at 298 K (determined by ¹H NMR; [H₂] was corrected for the presence of 25% of *para*-H₂, which is NMR-silent). As is often the case with excessively stable compounds, **3W** inhibits hydrosilylation. The origin of inhibition is readily traced to the presence of dihydrogen. Thus, reaction in a vial open to an inert atmosphere shows an almost threefold acceleration in initial turn-over frequency compared to the same reaction in a closed tube. Note that both samples have been maintained homogeneous to eliminate uncertainties of precipitation and have been taken from the same stock solution of acetophenone, HSiEt₃ and **1W**.

The key element of the catalyst activity and solubility at the final stages of the reaction of aliphatic substrates—the liquid clathrate—is metastable. It eludes characterization by rapidly changing its composition and converting into solid **2W** and **3W**. The liquid clathrates formed with aromatic substrates, on the other hand, are more stable and are amenable for analysis. Thus a liquid clathrate formed in the course of hydrosilylation of acetophenone has been characterized by ¹H NMR to have a composition of ~3.4 equivalents of the alkoxysilane Ph(Me)CH–OSiEt₃ per equivalent of tungsten. Note that this particular liquid clathrate is not critical to the catalyst activity, as the catalyst is somewhat soluble in the aromatic substrates anyway. In a broader sense, however, the characterization of this clathrate illustrates how only a few equivalents of the right component can retain the catalyst in a liquid phase, even if it is not soluble in the bulk of the reaction mixture²⁰.

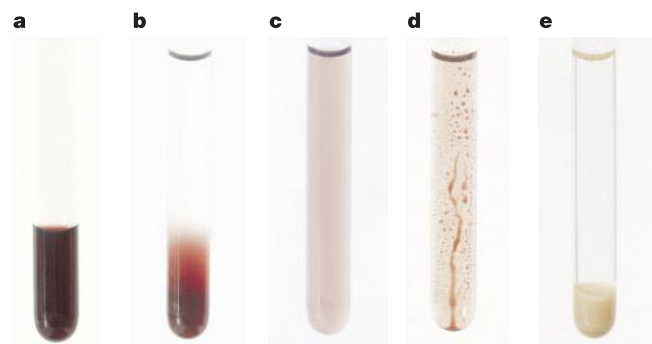


Figure 1 Photographs of the catalytic hydrosilylation of Et₂C=O by [CpW(CO)₂(IMes)]⁺[B(C₆F₅)₄][−]. **a**, Ketone complex **4W** before adding HSiEt₃. **b**, HSiEt₃ added, liquid not yet mixed. **c**, Mixed and homogeneous. **d**, Liquid clathrate formed. Reaction nearing completion. **e**, End of reaction. Catalyst has precipitated.

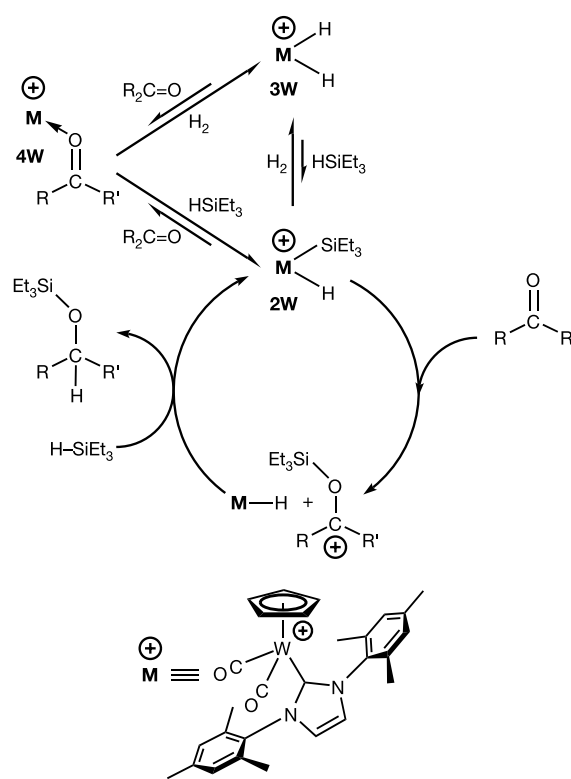


Figure 2 Proposed mechanism for catalytic ionic hydrosilylation.

The mechanism of hydrosilylation (Fig. 2) can be tentatively visualized as a modified ionic hydrogenation mechanism, which has been previously formulated for similar catalysts^{27,28}. Trialkylsilyl cation (silylium ion) is transferred from the metal to the ketone, yielding a carbocation intermediate. The carbocation abstracts a hydride (H^-) from $HSiEt_3$ or from the metal, furnishing alkoxy-silane and closing the catalytic cycle. The transfer of silylium and hydride ions could be concerted to some degree, but the observed cationic rearrangement in the reaction of cyclopropyl methyl ketone indicates an appreciable contribution of a cationic intermediate. A more detailed mechanistic investigation is needed to expand upon these preliminary findings.

The method for catalyst recycling and recovery presented here offers significant advantages, since it is solvent-free and requires no temperature changes, but this strategy is also prone to limitations. Examples of catalyst self-separation demonstrated thus far are restricted to aliphatic substrates, and the solvent-free approach is limited to both substrates being liquids. A challenge for future development is to tailor catalysts and hydrosilanes with the aim of engaging a broader scope of substrates into such solvent-free reactions with catalyst self-separation. A more general goal is to understand the fundamental aspects of liquid clathrate formation and to convert other catalysts into 'clathrate-enabled' catalysts, broadening the potential applicability of this recovery method to other classes of substrates and reactions. □

Received 11 April; accepted 23 June 2003; doi:10.1038/nature01856.

- Gladysz, J. A. Introduction: Recoverable catalysts and reagents—perspective and prospective. *Chem. Rev.* **102**, 3215–3216 (2002).
- Gladysz, J. A. Recoverable catalysts. Ultimate goals, criteria of evaluation, and the green chemistry interface. *Pure Appl. Chem.* **73**, 1319–1324 (2001).
- Cole-Hamilton, D. J. Homogeneous catalysis — new approaches to catalyst separation, recovery, and recycling. *Science* **299**, 1702–1706 (2003).
- Kragl, U. & Dwar, T. The development of new methods for the recycling of chiral catalysts. *Trends Biotechnol.* **19**, 442–449 (2001).
- Horváth, I. T. Fluorous biphasic chemistry. *Acc. Chem. Res.* **31**, 641–650 (1998).
- Gladysz, J. A. & Curran, D. P. Fluorous chemistry: From biphasic catalysis to a parallel chemical universe and beyond. *Tetrahedron* **58**, 3823–3825 (2002).
- Zuwei, X., Ning, Z., Yu, S. & Li, K. Reaction-controlled phase-transfer catalysis for propylene epoxidation to propylene oxide. *Science* **292**, 1139–1141 (2001).
- Houdin, G., Germain, A., Moreau, C. & Fajula, F. The catalysis of the Ruff oxidative degradation of aldonic acids by copper(II)-containing solids. *J. Catal.* **209**, 217–224 (2002).
- Wang, Y., Jiang, J., Zhang, R., Liu, X. & Jin, Z. Thermoregulated phase transfer ligands and catalysis IX. Hydroformylation of higher olefins in organic monophase catalytic system based on the concept of critical solution temperature of the nonionic tensioactive phosphine ligand. *J. Mol. Catal. A* **157**, 111–115 (2000).
- Wende, M., Meier, R. & Gladysz, J. A. Fluorous catalysis without fluorinated solvents. *J. Am. Chem. Soc.* **123**, 11490–11491 (2001).
- Wende, M. & Gladysz, J. A. Fluorous catalysis under homogeneous conditions without fluorinated solvents: A "greener" catalyst recycling protocol based upon temperature-dependent solubilities and liquid/solid phase separation. *J. Am. Chem. Soc.* **125**, 5861–5872 (2003).
- Ishihara, K., Kondo, S. & Yamamoto, H. 3,5-Bis(perfluorodecyl)phenylboronic acid as an easily recyclable direct amide condensation catalyst. *Synlett* 1371–1374 (2001).
- Ishihara, K., Hasegawa, A. & Yamamoto, H. A fluorinated super Bronsted acid catalyst: Application to fluorinated catalysis without fluorinated solvents. *Synlett* 1299–1301 (2002).
- Xiang, J. N., Orita, A. & Otera, J. Fluoroalkyldistannoxane catalysts for transesterification in fluorinated biphasic technology. *Adv. Synth. Catal.* **344**, 84–90 (2002).
- Atwood, J. L. In *Coordination Chemistry of Aluminum* (ed. Robinson, G. H.) 197–232 (VCH, New York, 1993).
- Steed, J. W. & Atwood, J. L. *Supramolecular Chemistry* 707 (Wiley, Chichester, 2002).
- Anastas, P. T. & Kirchhoff, M. M. Origins, current status, and future challenges of green chemistry. *Acc. Chem. Res.* **35**, 686–694 (2002).
- Anastas, P. T. & Zimmerman, J. B. Design through the 12 principles of green engineering. *Environ. Sci. Technol.* **37**, 94A–101A (2003).
- DeSimone, J. M. Practical approaches to green solvents. *Science* **297**, 799–803 (2002).
- Holbrey, J. D. et al. Liquid clathrate formation in ionic liquid–aromatic mixtures. *Chem. Commun.* 476–477 (2003).
- Lambert, J. B., Zhao, Y., Wu, H., Tse, W. C. & Kuhlmann, B. The allyl leaving group approach to tricoordinate silyl, germyl, and stannyl cations. *J. Am. Chem. Soc.* **121**, 5001–5008 (1999).
- Korolev, A. V., Ihara, E., Young, V. G. Jr & Jordan, R. F. Cationic aluminum alkyl complexes incorporating aminotroponimine ligands. *J. Am. Chem. Soc.* **123**, 8291–8309 (2001).
- Borovik, A. S. & Barron, A. R. Arene-mercury complexes stabilized by aluminum and gallium chloride: Catalysts for H/D exchange of aromatic compounds. *J. Am. Chem. Soc.* **124**, 3743–3748 (2002).
- Ojima, I. In *The Chemistry of Organic Silicon Compounds* (eds Patai, S. & Rappaport, Z.) 1479–1526 (Wiley, New York, 1989).
- Ojima, I., Li, Z. & Zhu, J. In *The Chemistry of Organic Silicon Compounds*, Vol. 2 (eds Rappaport, Z. & Apeloig, Y.) 1687–1792 (Wiley, New York, 1998).

- Dioumaev, V. K., Szalda, D. J., Hanson, J., Franz, J. A. & Bullock, R. M. An N-heterocyclic carbene as a bidentate hemilabile ligand: A synchrotron X-ray diffraction and density functional theory study. *Chem. Commun.* 1670–1671 (2003).
- Bullock, R. M. & Voges, M. H. Homogeneous catalysis with inexpensive metals: Ionic hydrogenation of ketones with molybdenum and tungsten catalysts. *J. Am. Chem. Soc.* **122**, 12594–12595 (2000).
- Voges, M. H. & Bullock, R. M. Catalytic ionic hydrogenations of ketones using molybdenum and tungsten complexes. *J. Chem. Soc. Dalton Trans.* 759–770 (2002).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements This research was supported by the Division of Chemical Sciences, Office of Basic Energy Sciences, of the US Department of Energy.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.B. (bullock@bnl.gov).

Southern Ocean origin for the resumption of Atlantic thermohaline circulation during deglaciation

Gregor Knorr^{1,2} & Gerrit Lohmann^{1,2}

¹Institut für Meteorologie, Universität Hamburg, Bundesstrasse 55, 20146 Hamburg, Germany

²Fachbereich für Geowissenschaften und Forschungszentrum Ozeanränder, Universität Bremen, PO 330 440, 28334 Bremen, Germany

During the two most recent deglaciations, the Southern Hemisphere warmed before Greenland^{1,2}. At the same time, the northern Atlantic Ocean was exposed to meltwater discharge³, which is generally assumed to reduce the formation of North Atlantic Deep Water^{4,5}. Yet during deglaciation, the Atlantic thermohaline circulation became more vigorous, in the transition from a weak glacial to a strong interglacial mode⁶. Here we use a three-dimensional ocean circulation model⁷ to investigate the impact of Southern Ocean warming and the associated sea-ice retreat⁸ on the Atlantic thermohaline circulation. We find that a gradual warming in the Southern Ocean during deglaciation induces an abrupt resumption of the interglacial mode of the thermohaline circulation, triggered by increased mass transport into the Atlantic Ocean via the warm (Indian Ocean) and cold (Pacific Ocean) water route^{9,10}. This effect prevails over the influence of meltwater discharge, which would oppose a strengthening of the thermohaline circulation. A Southern Ocean trigger for the transition into an interglacial mode of circulation provides a consistent picture of Southern and Northern hemispheric climate change at times of deglaciation, in agreement with the available proxy records.

Ice-core and ocean-sediment records reveal that during the last deglaciation, warming in the Southern Hemisphere preceded Greenland warming by more than 1,000 years (ref. 1), a time lag that was even longer for the penultimate deglaciation². One candidate for an interhemispheric teleconnection is provided by the oceanic thermohaline circulation (THC). Proxy data⁶ and modelling studies^{7,11,12} indicate that the last deglaciation was characterized by a transition from a weak glacial to a strong interglacial Atlantic THC. During the Bølling–Allerød (B/A) warm period, the sea surface temperatures (SSTs) in the North Atlantic almost reached interglacial values, consistent with active deep water formation, corroborated by benthic $\delta^{13}C$ data⁶. While most studies investigate processes responsible for a shut-down of the THC, we consider the question of how the conveyor gets restarted after a glacial mode with

weak overturning, using a numerical model for interglacial and glacial background conditions (see Methods section).

For the present-day state, we obtain an Atlantic Ocean circulation with export of 14 Sv ($1 \text{ Sv} = 10^6 \text{ m}^3 \text{ s}^{-1}$) at 30°S and maximum heat transport of 1.1 PW, which is in the range of observations¹³. For the Last Glacial Maximum (LGM), our modelled ocean circulation is characterized by a weaker circulation of about 8.5 Sv (Fig. 1a) and southward shifted convection sites (Fig. 1b), compared to present-day sites. North Atlantic Deep Water (NADW) is formed in the subpolar North Atlantic, and the North Atlantic Current flows in a zonal direction along the horizontal density gradients, consistent with reconstructions⁶.

To simulate the resumption of the Atlantic THC during deglaciation, we assume a linear transition from glacial to interglacial values in Southern Ocean temperature, sea ice and wind stress over 1,500 years, consistent with a deglacial sea-ice retreat within 1,000 years in the Southern Ocean south of the present-day polar front⁸. Caused by Southern Ocean warming and accompanying sea-ice retreat, a strong increase in overturning circulation is detected in experiment LGM_100 (Fig. 2). The warming generates a negative density anomaly in the Southern Ocean, inducing anomalous westward velocities between 30°S and 50°S , consistent with geostrophic balance (Fig. 3a). Off the Brazilian coast, the anomalous flow is along with pressure gradients to the north. Part of the relatively warm and salty water originating from the Indian Ocean is directly

injected into the subtropical South Atlantic via the eastern boundary current off the Namibian coast (Fig. 3). The thermal anomaly is attenuated along the northward conveyor route, but the saline characteristics of the warm water route persist¹⁴ (Fig. 2b). The salinity increase in the upper layer of the North Atlantic preconditions NADW formation in a fashion comparable to a mechanism⁹ that gradually intensifies convection during the first 1,000 years (Fig. 2b). The mechanism involves a positive feedback of THC owing to salinity advection¹⁵. Furthermore, sea-ice retreat in the Southern Ocean increases volume transport through the Drake Passage (25 Sv to 54 Sv) and the Atlantic import at 30°S of relatively cold and fresh water between 10°W and 50°W (Fig. 3).

Along with increased Atlantic overturning, the meridional heat transport warms the surface water of the North Atlantic, triggering an abrupt sea-ice retreat after 1,000 years. Besides the advective feedback, a convective feedback contributes to the resumption of the THC. Once convection is initiated in parts of the formerly ice-covered North Atlantic, the relatively warm and saline water masses from deeper layers coming up to the surface lose their heat readily, but keep their salt, and thus reinforce the starting process of convection (Fig. 2b). As a result, the mean convection depth in the Atlantic between 50°N and 60°N increases abruptly from 0 to

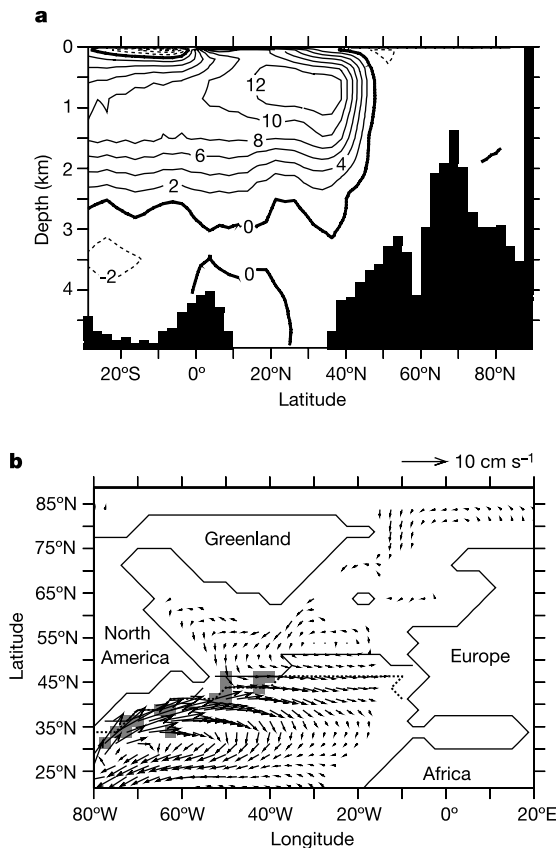


Figure 1 Modelled overturning streamfunction, as well as NADW convection sites and horizontal surface velocities in the Atlantic for the glacial control climate (LGM_CTRL). **a**, Meridional streamfunction (Sv); warm surface water flows northward to the subpolar regions, sinks, and flows southward as NADW, represented by the solid lines. **b**, Horizontal velocity (cm s^{-1}) in the upper 50 m of the North Atlantic. Major NADW convection sites are represented by the shaded area; summer and winter sea-ice margins are indicated by the solid and the dashed line, respectively.

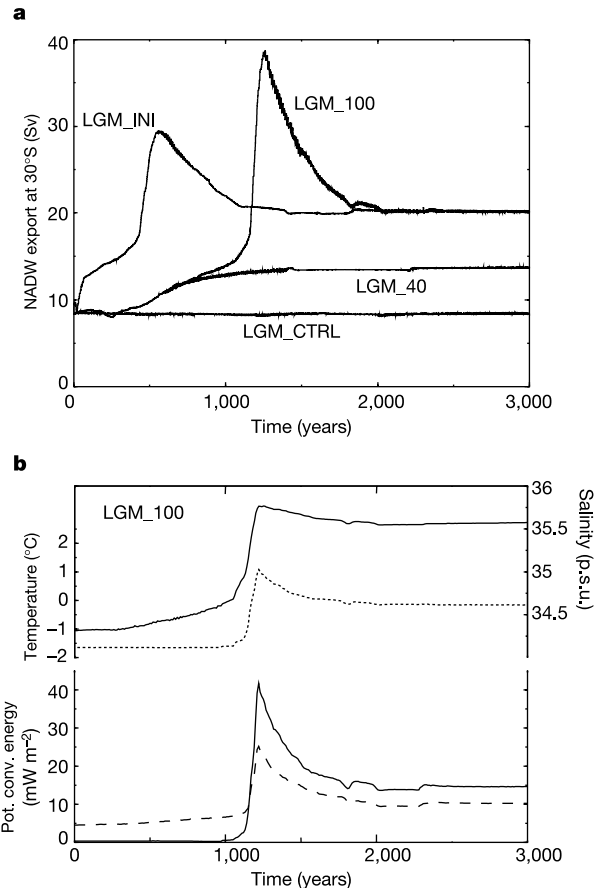


Figure 2 Temporal changes of NADW export at 30°S , and North Atlantic temperature, salinity and convection energy loss. **a**, In LGM_100, glacial conditions in the Southern Ocean (south of 30°S) are gradually replaced by interglacial conditions within 1,500 years. In LGM_40 gradual changes to interglacial conditions as in LGM_100 are assessed, but the transition terminates after 600 years, equivalent to a sea-ice retreat of 40%. Experiment LGM_INI represents instantaneously applied interglacial conditions of temperature, sea ice and wind stress. The curve LGM_CTRL shows the glacial control run. **b**, LGM_100 time series of SST ($^\circ \text{C}$; dotted curve), salinity (p.s.u.; solid curve) and potential energy loss by convection (mW m^{-2}) in the North Atlantic averaged between 55°N and 65°N (solid curve) and between 40°N and 55°N (dashed curve).

140 m. The convection in large parts of the Labrador Sea is only temporarily active, whereas the convection sites south of Iceland remain in operation (see the animations in Supplementary Information). The maximum heat transport in the Atlantic Ocean increases from 0.8 PW to 1.6 PW after 3,000 model years, causing an SST rise of more than 6 K in the North Atlantic (Fig. 3b).

To investigate the abrupt warming in more detail, we have run experiment LGM_INI, where background conditions in the Southern Ocean are instantaneously switched from glacial to interglacial ones. The model response (Fig. 2a) is similar to LGM_100, indicating that the abrupt warming is independent on the timescale of sea-ice margin retreat in the Southern Ocean. In order to estimate the threshold for the nonlinear transition shown in Fig. 2a, a series of experiments with different sea-ice retreat scenarios in the Southern Ocean were performed, showing that ~50% sea-ice margin retreat from glacial to interglacial conditions is sufficient to cause the abrupt model response. Experiment LGM_40, with 40% sea-ice margin retreat, is representative of the experiments with a linear model response (Fig. 2a). Despite enhanced NADW formation in this experiment, the sea-ice margin in the North Atlantic is almost unaffected, and the feedback loop for the nonlinear resumption of the THC described above is not active, suggesting that the level of Southern Ocean warming governs the

abrupt sea-ice retreat in the northern North Atlantic.

Furthermore, model simulations with alternated wind and temperature forcing reveal that imposed temperature and sea-ice changes in the Southern Hemisphere have a strong influence on THC strength compared to locally alternated wind fields. A further experiment showed that Northern Hemisphere warming has much less influence on NADW overturning strength than Southern Hemisphere warming (Supplementary Fig. 1).

In addition, we looked at the deglacial Heinrich sequence that is characterized by meltwater discharge to the North Atlantic³, whose destabilizing impact on THC has been proposed to suppress warming in the north¹⁶. Our glacial reference state (LGM_CTRL) is monostable with respect to transient North Atlantic freshwater pulses^{7,12}, as revealed by experiment LGM_CTRL_H500 (Fig. 4). This explains the observed recovery of the reduced glacial THC after Heinrich events⁶. Different experiments with transient (LGM_SH_H500) and permanent (LGM_SH_Hperm) freshwater flow to the North Atlantic (Fig. 4a) show that the abrupt resumption of Atlantic THC via Southern Ocean warming prevails over the destabilizing impact of meltwater discharge during the Heinrich sequence. Compared to LGM_100, the overturning response in LGM_SH_H500 and LGM_SH_Hperm is damped in magnitude and delayed in time.

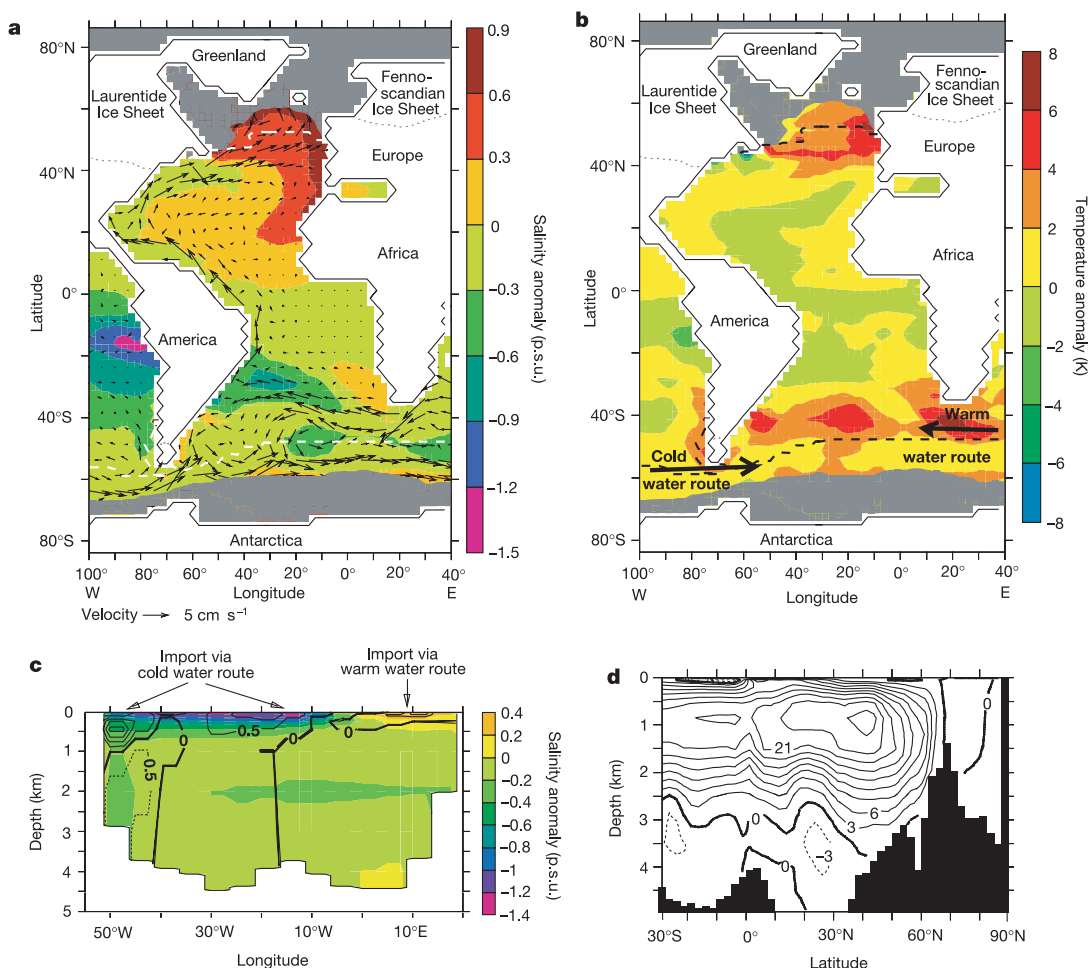


Figure 3 Differences between LGM_100 and LGM_CTRL, and meridional overturning streamfunction in LGM_100 after 3,000 model years. **a**, Annual mean salinity (p.s.u.) and horizontal velocity (cm s⁻¹) anomalies averaged over the upper 800 m. The grey shaded area and the dashed line show the annual mean 2 m height ice extent in experiment LGM_100 and LGM_CTRL, respectively. The maximum glacial continental ice margin is represented by the dotted lines. **b**, Annual mean temperature anomaly (K) averaged over

the upper 50 m; schematically overlaid with the main pathways of water import to the Atlantic Ocean. Sea-ice margins and continental ice sheets as in **a**. **c**, Annual mean salinity anomaly (p.s.u.) and anomaly of the meridional velocity component (cm s⁻¹) at 30° S. Solid lines indicate northward flow, dashed lines represent southward flow. **d**, Meridional overturning streamfunction (Sv) in the Atlantic in LGM_100 after 3,000 model years.

So far, our investigations reveal that Southern Ocean warming and associated sea-ice retreat enhances the transport via warm and cold water routes, causing a THC flow regime that is stronger and bistable (Fig. 4b). The bistable flow regime is driven by heat loss, whereas freshwater tends to reduce overturning strength¹⁵. The restarted THC bears the potential to react to deglacial meltwater input to the North Atlantic with a substantial long-term THC weakening, in accordance with the Younger Dryas cold sequence^{4,17} (Fig. 4b).

As a result of the restarted conveyor circulation (Fig. 3d), the maximum heat transport and the temperature in the North Atlantic increase dramatically (Fig. 3b), consistent with a temperature rise of more than 6 K for the B/A warm transition¹⁸. At the same time, temperatures in the tropical¹⁷ and South Atlantic¹⁹ decrease, in accordance with the oceanic interhemispheric teleconnection (increased THC cools the Southern Hemisphere²⁰). In our model

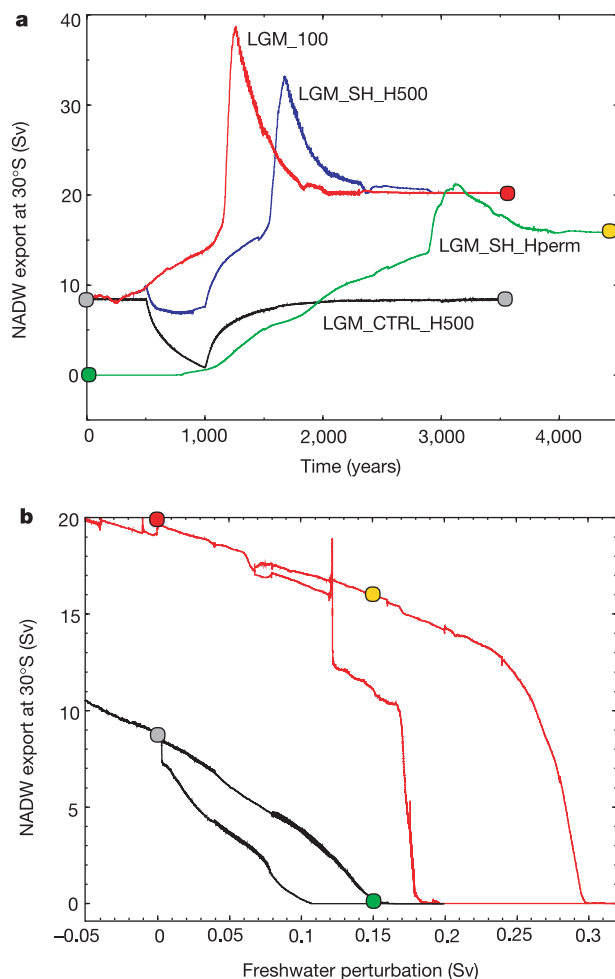


Figure 4 Atlantic NADW export at 30°S for the different freshwater pulse experiments and stability diagrams for the Atlantic THC. **a**, In LGM_CTRL_H500 and LGM_SH_H500, a transient freshwater pulse to the North Atlantic has been simulated between model year 500 and 1,000. In LGM_SH_perm, the freshwater input persists. Southern Ocean warming is applied in LGM_SH_H500 and LGM_SH_perm (also see Methods section). The coloured points indicate the different equilibrium climate states. **b**, The stability diagram (procedure explained in Methods section) of the glacial ocean circulation (black curve) differs substantially from that of LGM_100 (red curve). The restarted Atlantic THC exhibits a pronounced bistability for anomalous freshwater fluxes between +0.12 Sv and +0.3 Sv. The glacial THC shows a monostable behaviour at zero freshwater perturbation, and only weak bistability in the area of positive freshwater input. The location of climate states from **a** are displayed by the coloured points.

simulations, Southern Hemisphere cooling, as a response to a resumption of the THC in the North Atlantic, is inhibited by the experimental design. On the basis of the present investigation, the Antarctic Cold Reversal can be associated with an abrupt transition from a weak glacial to a strong interglacial ocean circulation mode that effectively transported heat to the North Atlantic. The corresponding B/A warm phase occurs only after a certain amount of heat has been transported to the North Atlantic, triggering a northern sea-ice retreat, which is then abruptly amplified by convective processes. We speculate that in conjunction with other effects^{21,22}, the increase in northward oceanic heat transport can contribute to the reduction of the great Northern Hemisphere ice sheets over North America and Scandinavia. The respective meltwater input weakens, but does not stop, NADW formation²³. The involved mechanism is consistent with the phase offsets between Southern and Northern hemisphere climate change on Milankovitch time-scales^{1,2}, and provides a possible explanation for the onset of the B/A. This mechanism represents an alternative explanation to the proposal that meltwater from Antarctica retracts Antarctic Bottom Water formation, leading to a switch-on of THC at a critical threshold¹⁶, an effect that would amplify the mechanism proposed here.

Our study suggests that the “Achilles Heel”¹⁰ and the “Flywheel” of the Atlantic overturning circulation are located in the North Atlantic and the Southern Ocean, respectively. Interestingly, the tropical foraminifer *Globorotalia menardii* is re-seeded in the Atlantic Ocean at the last glacial–interglacial transition²⁴. The only way to re-seed *G. menardii* in the Atlantic is around the Cape of Good Hope, leading to speculation²⁵ that the reopening of the Agulhas gap at the end of the last ice age may have played a role in restarting the Atlantic THC. Southern Ocean warming and associated sea-ice retreat might be the response to local Milankovitch forcing on the precessional period²⁶ or the result of tropical SST anomalies that transmitted from the tropical Pacific to the Antarctic region²⁷. Since the South Atlantic is characterized by a number of unique dynamical features, such as the large Agulhas Rings that form a key link in the THC^{9,28}, it is of interest to investigate the mechanism proposed here, using high-resolution models of the South Atlantic region to obtain a more detailed view in this region. More well-dated high-resolution records from terrestrial and marine sites in the Southern Hemisphere will be needed to decipher these additional pathways. They seem to be necessary for the prediction of future climate scenarios with respect to rising atmospheric CO₂ concentration and its impact on the Atlantic overturning circulation. □

Methods

Model description

The ocean model is based on the three-dimensional large-scale geostrophic model LSG². It includes a simple thermodynamic sea-ice model, a new advection scheme for temperature and salinity, and parameterization of overflow²⁹. The horizontal resolution is 3.5° on a semi-staggered grid (type ‘E’) with 11 levels in the vertical. For glacial conditions the storage of water in inland ice sheets is taken into account by setting all ocean points with present-day water depth less than 120 m to land. This causes the Bering Strait to be closed, and several shallow areas like the Arctic shelves to become land. To drive the ocean, monthly fields of wind stress, surface air temperature and freshwater flux are taken from a present-day and LGM simulation of the atmospheric general circulation model ECHAM3/T42³⁰. We use a hybrid coupled modelling approach, which allows an adjustment of surface temperatures and salinity to changes in the ocean circulation, based on an atmospheric energy balance model²⁹. The applied heat flux parameterization has been shown to be a suitable choice, allowing the simulation of observed SSTs and the maintenance of large-scale temperature anomalies in perturbation experiments²⁹. No flux correction is applied for both present-day and LGM boundary conditions.

Experimental design

In order to simulate the resumption of the Atlantic conveyor belt circulation after glacial times, we perform several experiments where interglacial values south of 30°S are applied as a surrogate for Southern Hemisphere warming. Each experiment is started from the glacial equilibrium climate and is integrated until a quasi steady state is reached. All experimental time series are smoothed by a one-year boxcar-average.

Freshwater pulse experiments

The experiments in Fig. 4a simulate deglacial meltwater discharge to the Atlantic Ocean between 20°N and 50°N. Experiment LGM_CTRL_H500 and LGM_SH_H500 display the model response to a 500-year freshwater-pulse of 0.15 Sv. The starting point of experiment LGM_SH_Hperm, representing the THC 'off-mode', results from a permanent freshwater flow of 0.15 Sv to the glacial equilibrium state. The freshwater flow persists during this experiment. In LGM_SH_H500 and LGM_SH_Hperm, Southern Hemisphere warming is executed as in LGM_100, while the glacial background conditions in experiment LGM_CTRL_H500 remain unaltered.

Stability analysis

We apply a slowly varying freshwater anomaly with a rate of $5 \times 10^{-5} \text{ Sv yr}^{-1}$ ($1 \times 10^{-6} \text{ Sv yr}^{-1}$ at the transition of the hysteresis curve that starts from LGM_100) uniformly between 20°N and 50°N to the Atlantic Ocean. Integration starts at the upper branches with zero freshwater forcing, which is then increased up to 0.2 Sv and 0.32 Sv for LGM_CTRL and LGM_100, respectively. The integration proceeds on the lower branch with freshwater input decreasing until -0.12 Sv. Then the freshwater increases again to close the loops. Owing to the slowly varying nature of the surface forcing the model is in quasi-equilibrium during the integration.

Received 21 November 2002; accepted 23 June 2003; doi:10.1038/nature01855.

1. Sowers, T. & Bender, M. Climate records covering the last deglaciation. *Science* **269**, 210–214 (1995).
2. Petit, R. *et al.* Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. *Nature* **399**, 429–436 (1999).
3. Marshall, J. S. & Clarke, G. K. C. Modelling North American freshwater runoff through the last glacial cycle. *Quat. Res.* **52**, 300–315 (1999).
4. Stocker, T. F. & Wright, D. G. Rapid transitions of the ocean's deep circulation induced by changes in surface water fluxes. *Nature* **351**, 729–732 (1991).
5. Maier-Reimer, E., Mikolajewicz, U. & Hasselmann, K. Mean circulation of the Hamburg LSG OGCM and its sensitivity to the thermohaline surface forcing. *J. Phys. Oceanogr.* **23**, 731–757 (1993).
6. Sarinthein, M. *et al.* Changes in east Atlantic deepwater circulation over the last 30,000 years: Eight time slice reconstructions. *Paleoceanography* **9**, 209–267 (1994).
7. Prange, M., Romanova, V. & Lohmann, G. The glacial thermohaline circulation: Stable or unstable? *Geophys. Res. Lett.* **29**, 10.1029/2002GL015337 (2002).
8. Shemesh, A. *et al.* Sequence of events during the last deglaciation in Southern Ocean sediments and Antarctic ice cores. *Paleoceanography* **17**, 101029/2000PA000599 (2002).
9. Gordon, A. L., Weiss, R. F., Smethie, W. M. Jr & Warner, M. J. Thermocline and intermediate water communication between the South Atlantic and Indian Oceans. *J. Geophys. Res.* **97**, 7223–7240 (1992).
10. Broecker, W. S. The great ocean conveyor. *Oceanography* **4**, 79–89 (1991).
11. Shin, S. I., Otto-Bliesner, B., Brady, E. C., Kutzbach, J. E. & Harrison, S. P. A simulation of the last glacial maximum climate using the NCAR-CCSM. *Clim. Dyn.* **20**, 127–151 (2003).
12. Ganopolski, A. & Rahmstorf, S. Rapid changes of glacial climate simulated in a coupled climate model. *Nature* **409**, 153–158 (2001).
13. MacDonald, A. M. & Wunsch, C. An estimate of global ocean circulation and heat fluxes. *Nature* **382**, 436–439 (1996).
14. Weijer, W., De Ruijter, W. P. M., Sterl, A. & Drijfhout, S. S. Response of the Atlantic overturning circulation to South Atlantic sources of buoyancy. *Glob. Planet. Change* **34**, 293–311 (2002).
15. Stommel, H. Thermohaline convection with two stable regimes of flow. *Tellus* **13**, 224–230 (1961).
16. Stocker, T. F. in *Continuum Mechanics and Applications in Geophysics and the Environment* (eds Straughan, B., Greeve, R., Ehrentauf, H. & Wang, Y.) 337–367 (Springer, New York, 2001).
17. Rühlemann, C., Mulitza, S., Müller, P. J., Wefer, G. & Zahn, R. Warming of the tropical Atlantic Ocean and slowdown of thermohaline circulation during the last deglaciation. *Nature* **402**, 511–514 (1999).
18. Bard, E., Rostek, F., Turon, J. L. & Gendreau, S. Hydrological impact of Heinrich events in the subtropical northeast Atlantic. *Science* **289**, 1321–1324 (2000).
19. Sachs, J. P., Anderson, R. F. & Lehman, S. J. Glacial surface temperatures of the southeast Atlantic Ocean. *Science* **293**, 2077–2079 (2001).
20. Crowley, T. J. North Atlantic Deep Water cools the Southern Hemisphere. *Paleoceanography* **7**, 489–497 (1992).
21. Stephens, B. B. & Keeling, R. F. The influence of Antarctic sea ice on glacial–interglacial CO₂ variations. *Nature* **404**, 171–174 (2000).
22. Toggweiler, J. R. Variation of atmospheric CO₂ by ventilation of the ocean's deepest water. *Paleoceanography* **14**, 571–588 (1999).
23. Lohmann, G. & Schulz, M. Reconciling Bolling warmth with peak deglacial meltwater discharge. *Paleoceanography* **15**, 537–540 (2000).
24. Schott, W. Die Foraminiferen in dem äquatorialen Teil des Atlantischen Ozeans. *Deut. Atl. Exped. Meteor* 1925–1927 **3**, 43–134 (1935).
25. Berger, W. H. & Wefer, G. in *The South Atlantic: Present and Past Circulation* (eds Wefer, G., Berger, W. H., Siedler, G. & Webb, D. J.) 363–410 (Springer, Heidelberg, 1996).
26. Kim, S. J., Crowley, T. J. & Stössel, A. Local orbital forcing of Antarctic climate change during the last interglacial. *Science* **280**, 728–730 (1998).
27. Koutavas, A., Lynch-Stieglitz, J., Marchitto, T. M. Jr & Sachs, J. P. El Niño-like pattern in ice age tropical Pacific sea surface temperature. *Science* **297**, 226–230 (2002).
28. Schouten, M. W., De Ruijter, P. M. & van Leeuwen, P. J. Upstream control of Agulhas Ring shedding. *J. Geophys. Res.* **107** 101029/2001JC000804 (2002).
29. Prange, M., Lohmann, G. & Paul, A. Influence of vertical mixing on the thermohaline hysteresis: Analyses of an OGCM. *J. Phys. Oceanogr.* **33**, 1707–1721 (2003).
30. Lohmann, G. & Lorenz, S. The hydrological cycle under paleoclimatic boundary conditions as derived from AGCM simulations. *J. Geophys. Res.* **105**, 17417–17436 (2000).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank H. Jansen, S. Mulitza and M. Prange for suggestions. L. Könnecke, S. Schubert, M. Butzin and S. Blessing are acknowledged for their technical support. This work was supported by BMBF through the DEKLIM project 'climate transitions'.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to G.K. (gregor.knorr@dkrz.de).

Possible thermal and chemical stabilization of body-centred-cubic iron in the Earth's core

Lidunka Vočadlo¹, Dario Alfè^{1,2}, M. J. Gillan², I. G. Wood¹, J. P. Brodholt¹ & G. David Price¹

¹Research School of Earth Sciences, Birkbeck College and University College London, and

²Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK

The nature of the stable phase of iron in the Earth's solid inner core is still highly controversial. Laboratory experiments¹ suggest the possibility of an uncharacterized phase transformation in iron at core conditions and seismological observations^{2–4} have indicated the possible presence of complex, inner-core layering. Theoretical studies^{5,6} currently suggest that the hexagonal close packed (h.c.p.) phase of iron is stable at core pressures and that the body centred cubic (b.c.c.) phase of iron becomes elastically unstable at high pressure. In other h.c.p. metals, however, a high-pressure b.c.c. form has been found to become stabilized at high temperature. We report here a quantum mechanical study of b.c.c.-iron able to model its behaviour at core temperatures as well as pressures, using *ab initio* molecular dynamics free-energy calculations. We find that b.c.c.-iron indeed becomes entropically stabilized at core temperatures, but in its pure state h.c.p.-iron still remains thermodynamically more favourable. The inner core, however, is not pure iron, and our calculations indicate that the b.c.c. phase will be stabilized with respect to the h.c.p. phase by sulphur or silicon impurities in the core. Consequently, a b.c.c.-structured alloy may be a strong candidate for explaining the observed seismic complexity of the inner core^{2–4}.

Seismic measurements show that the Earth's solid inner core is complex. The detailed interpretations of these data differ, but all workers^{2–4} conclude that the inner core exhibits a significant degree of layering, which may either reflect the changing history of core crystallization, or the occurrence of an unidentified change in the core-forming phase. It is currently generally accepted that the core consists predominantly of Fe, and conventional interpretations of the elastic anisotropy of the inner core have been based on the idea of a partial alignment of crystals of h.c.p.-Fe⁵. Recently, however this interpretation has been challenged³, as has the assumption that h.c.p.-Fe is the thermodynamically stable polymorph of Fe at the high temperatures found in the inner core¹. The low pressure (*P*)–temperature (*T*) phase diagram of Fe is not controversial, with b.c.c.-Fe stable at ambient conditions, h.c.p.-Fe stable above about 15 GPa at low temperatures and the face-centred-cubic (f.c.c.) structure being stable above about 1,300 K at low pressures. High-*P/T* diamond-anvil-cell experiments⁷ suggest a new phase above ~40 GPa and ~1,000 K, but the structure of this phase is still unresolved. In addition, a solid–solid phase transition at around

~200 GPa and ~4,000 K has been inferred from shock data¹.

In spite of the arguments in favour of h.c.p.-Fe, it has been proposed by a number of workers that b.c.c.-Fe might be the stable high-*P/T* phase^{8,9}. A strong argument in favour of the stability of b.c.c.-Fe under these conditions is that a number of transition metals are known to transform from close-packed structures to the b.c.c. structure at temperatures just below their melting curve^{10,11}. For Fe, the argument in favour of b.c.c. is further strengthened by *ab initio* calculations showing the existence of an appreciable magnetic moment in b.c.c.-Fe at Earth's core pressures¹², which implies an additional magnetic entropy stabilization at high temperatures. However, for a long time it appeared that theoretical calculations had ruled out the b.c.c. structure as a candidate for the stable phase of Fe in the core. This was for three reasons: (1) *ab initio* calculations have shown that the structure becomes elastically unstable at pressures above 150 GPa^{12,13}; (2) *ab initio* calculations also reveal a high-pressure vibrational instability away from the zone centre¹⁴; (3) calculations show that the enthalpy of the perfect b.c.c. structure is considerably higher than that of the h.c.p. phase^{6,12–14}. However, these arguments are not conclusive because they are based on athermal or lattice dynamical^{6,12–14} *ab initio* calculations, which, because of the dynamical instability of b.c.c.-Fe, cannot be used to determine entropic effects at high pressures. In the past, computing limitations prevented a more sophisticated analysis, but recent methodological developments mean that it is now possible accurately to address these thermal effects. Thus, to resolve the controversy over the effect of temperature on the stability of b.c.c.-Fe, we report here *ab initio* molecular dynamics (AIMD) calculations on b.c.c.-Fe to simulate directly its behaviour at the high temperatures relevant to the Earth's inner core.

It is well established that *ab initio* calculations give an accurate description of all the key properties of Fe^{12–15}. The calculations presented here are based on density functional theory (DFT)¹⁶ within the generalized gradient approximation (GGA)¹⁷. They were performed using the VASP code¹⁸, using the projected-

augmented wave (PAW) method¹⁹ to calculate the total energy of the system. The PAW method is closely related to the ultrasoft pseudopotential method and has been shown to give results that agree accurately with all-electron methods^{20,21}. In AIMD, the ions are treated as classical particles and, for each set of atomic positions, the electronic energy and forces on the ions are calculated within the framework of DFT, which includes the thermal excitations of the electrons.

The elastic and vibrational instabilities of high-pressure b.c.c.-Fe have been studied before^{12–14}, but because of their key role in this work, we present in Fig. 1 our own results for the phonon dispersion relations as a function of pressure. It can be seen that the onset of the instability of b.c.c.-Fe at high pressure occurs below ~180 GPa where one of the branches in the X [110] direction becomes unstable at the zone centre (associated with the elastic constant $c' = (c_{11} - c_{12})/2$ becoming negative). This instability is associated with the b.c.c. → f.c.c. transition¹³. However, there is also a second instability occurring below 260 GPa, along the [111] direction (Γ to L); the pattern of the displacement involved in this instability concerns the relative motion of planes of atoms in the [111] direction of the cubic cell. This pattern of displacement is directly involved in the b.c.c. → ω transition, a transition which is known to readily occur in other transition metals^{10,11,22}.

In order to establish whether the b.c.c. phase at high pressure is entropically stabilized by temperature, we performed AIMD at 6,000 K (approximate core temperatures^{23,24}) for a chosen volume of 7.2 Å³ per atom (approximate core density) and compared the results with those from simulations performed at lower temperatures. The stability of the b.c.c. structure at high temperatures was assessed in four separate ways: (1) the anisotropic stresses were analysed for evidence of elastic stability, (2) the atomic positions with respect to those in a perfect b.c.c. structure were analysed for evidence of vibrational stability, (3) the structure factors were calculated to help identify the phases present, (4) the relative free energy with respect to the h.c.p. structure was calculated in order to

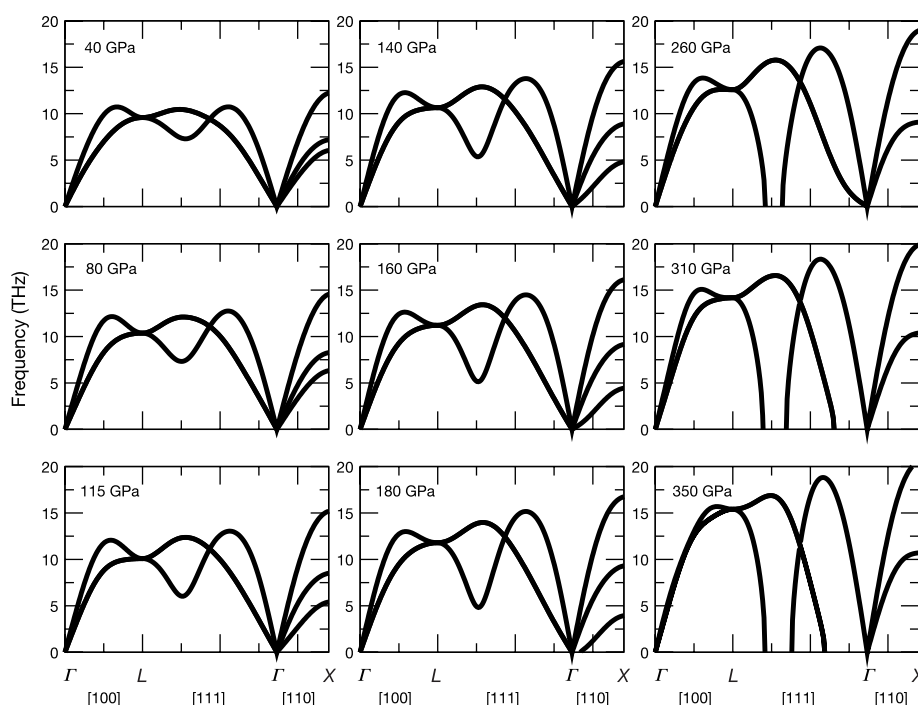


Figure 1 The calculated phonon dispersion curves for b.c.c.-Fe at nine different volumes (corresponding to ~40–350 GPa). The softening in the [110] direction at ~180 GPa indicates the onset of the b.c.c. → f.c.c. transition, and that in the [111] direction above ~260 GPa corresponds to the b.c.c. → ω transition. For these calculations, we used a

64-atom ($4 \times 4 \times 4$) cubic supercell with a $3 \times 3 \times 3$ *k*-point sampling grid and the small-displacement method to obtain the vibrational frequencies, the methodology of which has been described elsewhere²¹.

determine its thermodynamic stability at core conditions. The graph shown in the upper left corner of Fig. 2 gives the six components of the stress tensor of the cubic b.c.c. cell at 6,000 K as a function of simulation time; it can be seen that the supercell experiences hydrostatic stresses throughout the simulation, suggesting that the cell is stable. It is clear that when the temperature is considerably reduced (to below 3,000 K), the cell is no longer under hydrostatic stress, indicating that the b.c.c.-Fe structure is no longer elastically stable.

To identify whether the atoms in our simulation cell have moved away from the perfect b.c.c. structure, we can look at the positions of the atoms throughout the simulation relative to b.c.c. lattice sites. A convenient method for analysing the evolution of the atomic positions over the length of the simulation is to use the position correlation function, $p(t)$, which not only tells us where the atoms are, but also may indicate the existence of an imminent phase transition. The graph in the lower left corner of Fig. 2 shows $p(t)$ at 6,000 K; it is clear that $p(t) \rightarrow 0$ as $t \rightarrow \infty$, indicating that the atomic positions in the simulation cell are, on average, those of a perfect b.c.c. structure. We can therefore conclude that, at 6,000 K and a volume of $7.2 \text{ \AA}^3$ per atom (approximate core pressures and temperatures), the b.c.c. structure is vibrationally stable. However, as the system is cooled, it can readily be seen that the atoms deviate significantly and permanently away from a b.c.c. structure (that is, $p(t) \neq 0$ as $t \rightarrow \infty$) below 3,000 K. In fact, analysis of the structure factors (not presented) show that the b.c.c. phase undergoes a transition to the ω phase.

Having established that the b.c.c. phase under pressure is elasti-

cally and vibrationally stable at high temperatures, we need to calculate its free energy at core conditions to determine its thermodynamic stability with respect to h.c.p. We use the method of thermodynamic integration, which allows us to calculate the difference in free energy, $F - F_0$, between our *ab initio* b.c.c. system and a reference system whose potential energies are U and U_0 respectively. Recently^{21,23–25} we have successfully used this method to calculate *ab initio* the melting behaviour of Al at pressure, and the properties of liquid and h.c.p.-Fe. An important point to note is that these calculations were performed without spin polarization. As mentioned earlier, it is well known that the b.c.c. phase has a significant magnetic moment at core pressures¹², which, if maintained at high temperature, would contribute to the free energy through the disordering of the orientations of the spins. However, at high temperatures, the magnetic moment will be destroyed by thermal excitation of the electrons. We performed calculations on the b.c.c. structure at inner core pressures as a function of electronic temperature (2,000–6,000 K). Our results show that, as expected, the magnetic moment decreases with temperature, and disappears altogether well before core temperatures are reached; there is, therefore, no contribution to the free energy of b.c.c.-Fe from magnetic entropy.

Table 1 shows our calculated Helmholtz free energies for the b.c.c. and h.c.p. phases over a range of volumes and temperatures along (and below) our previously determined melting curve of h.c.p.-Fe^{23–24}. It is clear that in all cases $F_{\text{b.c.c.}} > F_{\text{h.c.p.}}$. However, the differences in free energies are small (33–58 meV per atom along the melting temperature curve, T_m). The Earth's inner core, however, is

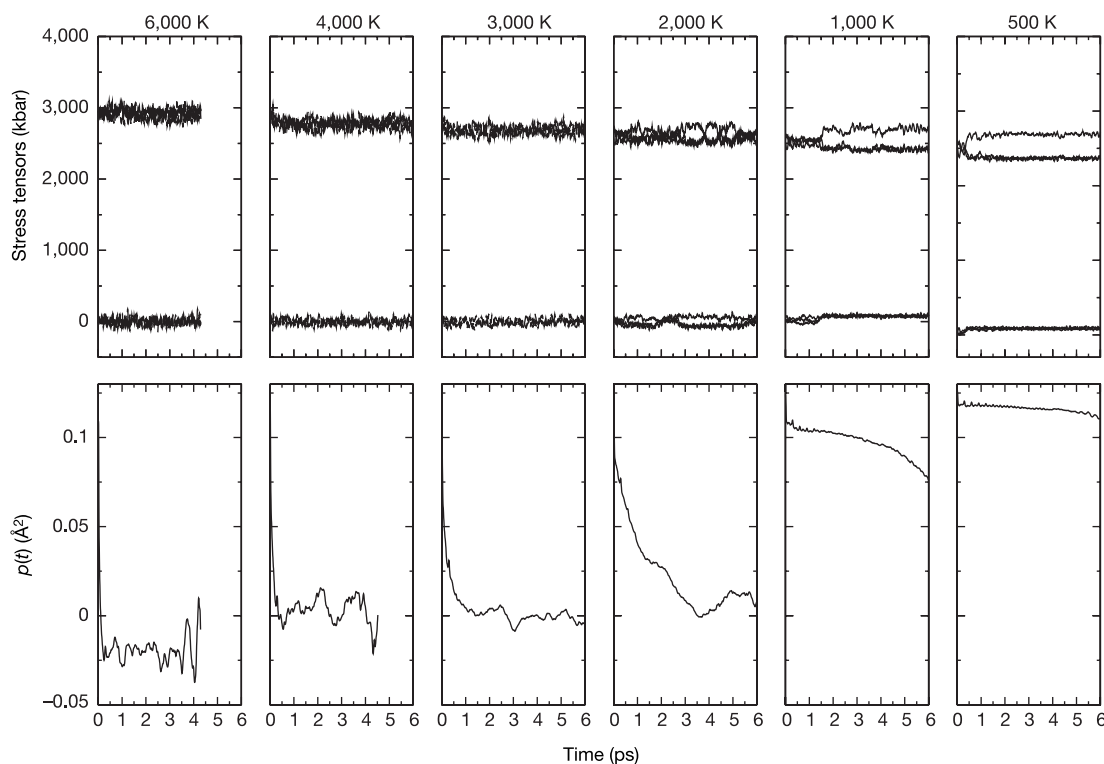


Figure 2 The calculated stress tensors as a function of simulation time (upper row) and position correlation functions (lower row) for a 64-atom cubic supercell at different temperatures. The structure deviates away from the b.c.c. phase below 3,000 K. We performed these high-temperature calculations on two supercells to ensure that any transition observed is not influenced by the size and shape of the simulation cell. We used both the 64-atom cubic supercell and a hexagonal supercell ($3 \times 3 \times 5$) of 135 atoms, since this latter setting provides us with a simulation cell which readily allows us to have within it an integral number of unit cells of both the b.c.c. and ω phases. The calculations were run initially at 6,000 K and then at successively lower temperatures, using a stable

equilibrated high-temperature b.c.c. configuration as the starting configuration in each case. The position correlation function for a chosen atom, i , is given by $p(t) = \langle (\mathbf{r}_i(t + t_0) - \mathbf{R}_i^0) \cdot (\mathbf{r}_i(t_0) - \mathbf{R}_i^0) \rangle$, where $\mathbf{r}_i$ is the time-varying position of the atom and $\mathbf{R}_i^0$ is the position of that atom's lattice site in the perfect b.c.c. structure. The angular brackets denote the thermal average, which in practice is evaluated as an average over time origins, t_0 , and atoms i . For long times t , vibrational displacements become uncorrelated, so that $p(t) = \langle (\mathbf{r}_i(t + t_0) - \mathbf{R}_i^0) \cdot (\mathbf{r}_i(t_0) - \mathbf{R}_i^0) \rangle \rightarrow \langle \mathbf{r}_i - \mathbf{R}_i^0 \rangle^2$, and if all atoms vibrate about b.c.c. lattice sites, $\langle \mathbf{r}_i - \mathbf{R}_i^0 \rangle = 0$, so that $p(t) \rightarrow 0$ as $t \rightarrow \infty$.

Table 1 Calculated Helmholtz free energy of the b.c.c. and h.c.p. phases of Fe

Volume ($\text{\AA}^3$)	Temperature (K)	$F_{\text{b.c.c.}}$ (eV)	$F_{\text{h.c.p.}}$ (eV)	$\Delta F_{\text{b.c.c.-h.c.p.}}$ (meV)
9.0	3,500	-10.063	-10.109	46
8.5	3,500	- 9.738	- 9.796	58
7.8	5,000	-10.512	-10.562	50
7.2	6,000	-10.633	-10.668	35
6.9	6,500	-10.545	-10.582	37
6.7	6,700	-10.288	-10.321	33
*7.2	3,000	- 7.757	- 7.932	175

Shown is the *ab initio* Helmholtz free energy per atom of the b.c.c. and h.c.p. phases of Fe, $F_{\text{b.c.c.}}$ and $F_{\text{h.c.p.}}$, at state points along (*and below) the calculated melting curve²³. In this work, the reference system is a simple inverse power potential which takes the form $U = 4\epsilon(T/r)^{\alpha}$ where $\epsilon = 1 \text{ eV}$, $T = 1.77 \text{ \AA}$, $\alpha = 5.86$. We have previously shown that this reference potential, based on only a repulsive term, describes the *ab initio* system extremely well; the bonding term is almost independent of the atomic positions, depending only on the volume and temperature of the system^{21,24}. We stress, however, that the final result is totally independent of the choice of reference system. We performed these calculations on a 64-atom supercell ($4 \times 4 \times 4$ primitive cells) with a $3 \times 3 \times 3$ *k*-point grid. Considerable effort was spent on convergence tests in *k*-point sampling to reduce the error in free energies to $< 10 \text{ meV}$ per atom. Cell size effects have been extensively studied in our previous work on h.c.p.-Fe²¹.

known not to be made of pure Fe, but is expected to be alloyed with between 5 to 10 mol.% of a lighter element. It is now reasonably well established that either separately or together S and Si are two of the most probable light elements alloyed with Fe in the core^{26–29}. Additionally, recent experiments³⁰ have shown that at high pressures FeSi crystallizes with the CsCl-structure (that is, has identical atomic coordinates to b.c.c.-Fe), and at lower concentrations Si stabilizes b.c.c.-Fe under pressure and temperature conditions at which pure Fe is found with the h.c.p. structure²⁶. We therefore investigated the energetic effect of the substitution of S and Si in b.c.c. and h.c.p.-Fe at representative core cell volumes ($7.2 \text{ \AA}^3$). We find that the enthalpies of the S and Si defects are respectively 1.4 and 1.2 eV per defect atom more stable in the b.c.c. structure than in the h.c.p. phase. Therefore, for example, a 5 mol.% concentration of Si in Fe, would stabilize the b.c.c. phase by 60 meV. Thus, in contrast with lower-pressure experiments²⁶, we conclude that the presence of S or Si as the light impurity element in the core, at appropriate concentrations, could favour the formation of a b.c.c.- rather than an h.c.p.-structured Fe alloy phase at temperatures just below T_m at inner core pressure.

In contrast to earlier work, we have shown that a b.c.c.-structured phase is entropically stabilized under inner core conditions, and is made more stable than the h.c.p.-phase by impurities in the inner core. This means that a b.c.c.-structured alloy is a strong candidate for explaining the recently observed^{2–4} seismic complexity of the inner core. □

Received 21 August 2002; accepted 18 June 2003; doi:10.1038/nature01829.

1. Brown, M. J. The equation of state of iron to 450 GPa: Another high-pressure solid phase? *Geophys. Res. Lett.* **28**, 4339–4342 (2001).
2. Song, X. D. & Helmerger, D. V. Seismic evidence for an inner core transition zone. *Science* **282**, 924–927 (1998).
3. Beghein, C. & Trampert, J. Robust normal mode constraints on inner-core anisotropy from model space search. *Science* **299**, 552–555 (2003).
4. Ishii, M. & Dziewonski, A. M. The innermost inner core of the earth: Evidence for a change in anisotropic behavior at the radius of about 300 km. *Proc. Natl Acad. Sci. USA* **99**, 14026–14030 (2002).
5. Steinle-Neumann, G., Stixrude, L., Cohen, R. E. & Gölseren, O. Elasticity of iron at the temperature of the Earth's inner core. *Nature* **413**, 57–60 (2001).
6. Vočadlo, L., Brodholt, J., Alfé, D., Gillan, M. J. & Price, G. D. *Ab initio* free energy calculations on the polymorphs of iron at core conditions. *Phys. Earth Planet. Inter.* **117**, 123–137 (2000).
7. Shen, G. Y., Mao, H. K., Hemley, R. J., Duffy, T. S. & Rivers, M. L. Melting and crystal structure of iron at high pressures and temperatures. *Geophys. Res. Lett.* **25**, 373–376 (1998).
8. Ross, M., Young, D. A. & Grover, R. Theory of the iron phase diagram at Earth core conditions. *J. Geophys. Res.* **95**, 21713–21716 (1990).
9. Matsui, M. & Anderson, O. L. The case for a body-centred-cubic phase (α') for iron at inner core conditions. *Phys. Earth Planet. Inter.* **103**, 55–62 (1997).
10. Petry, W. Dynamical precursors of martensitic transitions. *J. Phys. IV* **5**, C2-15–C2-28 (1995).
11. Trampenau, J. *et al.* Phonon dispersion of the bcc phase of group-IV metals. III. bcc hafnium. *Phys. Rev. B* **43**, 10963–10969 (1991).
12. Söderlind, P., Moriarty, J. A. & Willis, J. M. First-principles theory of iron up to Earth-core pressures: structural, vibrational and elastic properties. *Phys. Rev. B* **53**, 14063–14072 (1996).
13. Stixrude, L. & Cohen, R. E. Constraints on the crystalline structure of the inner core: mechanical instability of BCC iron at high pressures. *Geophys. Res. Lett.* **22**, 125–128 (1995).

14. Moriarty, J. A. in *High Pressure Science and Technology 1993* (eds Schmidt, S. C., Shaner, J. W., Samara, G. A. & Ross, M.) 233–236 (AIP Press, New York, 1994).
15. Mao, H. K. *et al.* Phonon density of states of iron up to 153 gigapascals. *Science* **292**, 914–916 (2001).
16. Hohenberg, P. & Kohn, W. Inhomogeneous electron gas. *Phys. Rev.* **136**, B864–B871 (1964).
17. Wang, Y. & Perdew, J. Correlation hole of the spin-polarized electron gas, with exact small-wave-vector and high-density scaling. *Phys. Rev. B* **44**, 13298–13307 (1991).
18. Kresse, G. & Furthmüller, J. Efficient iterative schemes for *ab initio* total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169–11186 (1996).
19. Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **50**, 17953–17979 (1994).
20. Alfé, D., Kresse, G. & Gillan, M. J. Structure and dynamics of liquid iron under Earth's core conditions. *Phys. Rev. B* **61**, 132–142 (2000).
21. Alfé, D., Price, G. D. & Gillan, M. J. Thermodynamics of hexagonal-close-packed iron under Earth's core conditions. *Phys. Rev. B* **64**, 045123 (2001).
22. Grad, G. B. *et al.* Electronic structure and chemical bonding effects upon the bcc to Ω phase transition: *ab initio* study of Y, Zr, Nb and Mo. *Phys. Rev. B* **62**, 12743–12753 (2000).
23. Alfé, D., Gillan, M. J. & Price, G. D. The melting curve of iron at Earth's core pressures from *ab initio* calculations. *Nature* **401**, 462–464 (1999).
24. Alfé, D., Price, G. D. & Gillan, M. J. Iron under Earth's core conditions: liquid-state thermodynamics and high pressure melting curve from *ab initio* calculations. *Phys. Rev. B* **65**, 165118 (2002).
25. Vočadlo, L. & Alfé, D. *Ab initio* melting curve of the fcc phase of aluminium. *Phys. Rev. B* **65**, 214105 (2002).
26. Lin, J.-F., Heinz, D. L., Campbell, A. J., Devine, J. M. & Shen, G. Iron-silicon alloy in Earth's core? *Science* **295**, 313–315 (2002).
27. Alfé, D., Price, G. D. & Gillan, M. J. Composition and temperature of the Earth's core constrained by combining *ab initio* calculations and seismic data. *Earth Planet. Sci. Lett.* **195**, 91–98 (2002).
28. Alfé, D., Price, G. D. & Gillan, M. J. *Ab initio* chemical potentials of solid and liquid solutions and the chemistry of the Earth's core. *J. Chem. Phys.* **116**, 7127–7136 (2002).
29. Alfé, D., Gillan, M. J. & Price, G. D. Constraints on the composition of the Earth's core from *ab-initio* calculations. *Nature* **405**, 172–175 (2000).
30. Dobson, D. P., Vočadlo, L. & Wood, I. G. A new high-pressure phase of FeSi. *Am. Mineral.* **87**, 784–787 (2002).

Acknowledgements L.V. and D.A. thank the Royal Society for their continued support through the University Fellowship scheme. We also thank NERC for providing computing facilities via grants.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to L.V. (l.vocadlo@ucl.ac.uk).

Measuring fast neutrons in Hiroshima at distances relevant to atomic-bomb survivors

T. Straume^{1,2}, G. Rugel^{3,4}, A. A. Marchetti², W. Rühm⁴, G. Korschinek³, J. E. McAninch², K. Carroll², S. Egbert⁵, T. Faestermann³, K. Knie³, R. Martinelli², A. Wallner^{3,4} & C. Wallner³

¹University of Utah, 729 Arapsee Drive, Suite 2334, Salt Lake City, Utah 84108, USA

²Lawrence Livermore National Laboratory (LLNL), PO Box 808, Livermore, California 94550, USA

³Technische Universität München, D-85747 Garching, Germany

⁴Ludwig Maximilians Universität München, D-80336, München, Germany

⁵Science Applications International Corporation, 10260 Campus Point Drive, San Diego, California 92121, USA

Data from the survivors of the atomic bombs serve as the major basis for risk calculations of radiation-induced cancer in humans¹. A controversy has existed for almost two decades, however, concerning the possibility that neutron doses in Hiroshima may have been much larger than estimated. This controversy was based on measurements of radioisotopes activated by thermal neutrons that suggested much higher fluences at larger distances than expected^{2–6}. For fast neutrons, which contributed almost all the neutron dose, clear measurement validation has so far proved impossible at the large distances (900 to 1,500 m) most relevant to survivor locations⁶. Here, the first

results are reported for the detection of ^{63}Ni produced predominantly by fast neutrons (above about 1 MeV) in copper samples from Hiroshima. This breakthrough was made possible by the development of chemical extraction methods^{7,8} and major improvements in the sensitivity of accelerator mass spectrometry for detection of ^{63}Ni atoms (refs 8–11). When results are compared with ^{63}Ni activation predicted by neutron doses for Hiroshima survivors⁶, good agreement is observed at the distances most relevant to survivor data. These findings provide, for the first time, clear measurement validation of the neutron doses to survivors in Hiroshima.

It was recognized in the final report of the 1986 reassessment of the atomic bomb radiation dosimetry for Hiroshima and Nagasaki (DS86)⁶ that the calculated neutron doses for survivors could possibly be wrong. The principal reasons for this concern were: (1) the observation that a large discrepancy was suggested between the available ^{60}Co activation measured in Hiroshima¹² and calculations based on DS86 and (2) the paucity of neutron validation measurements available at that time, preventing adequate resolution of this matter. It was concluded that the neutron doses are in doubt until further work is done and that special efforts should be made to provide additional neutron measurements⁶. This generated extensive interest and resulted in a large number of low-energy (thermal) neutron activation measurements in Hiroshima^{2–5,13–18} and to a lesser extent in Nagasaki¹⁹. See the review of ref. 20 for a more complete list of thermal neutron references. Even with the additional thermal neutron measurements, a definitive determination of the neutron dose has not been possible because of factors such as uncertainties in the measured background and uncertainties in the relationships between thermal neutrons and neutron dose. In contrast, fast neutrons produced essentially all of the neutron dose in Hiroshima and are much less dependent on sample environment. Although fast neutron measurements (above about 2 MeV) were made in Hiroshima within weeks of the bombings by beta-ray counting of the radioisotope ^{32}P (half-life, 14.2 days) in sulphur samples^{21–22}, the results reached background levels²³ at about 700 m from the hypocentre (the point on the ground below the explosion). Hence, they could not be used for direct validation of fast neutrons at the larger distances most relevant for the data on atomic-bomb survivors. For these reasons, it is clear that fast-neutron activation measurements at relevant distances in Hiroshima would be indispensable for validation of the neutron doses received by atomic-bomb survivors.

We have developed the capability to detect trace amounts of

^{63}Ni (half-life, $100.1 \pm 2.0 \text{ yr}^{24}$) in copper using accelerator mass spectrometry (AMS)^{7–11,25,26}. The amount of ^{63}Ni produced via the reaction $^{63}\text{Cu}(\text{n,p})^{63}\text{Ni}$ in copper samples exposed to atomic-bomb neutrons is a measure of the fast-neutron fluence^{25–27}. On the basis of the available cross-sections, 95% of the activation from this reaction was by neutrons above 1.5 MeV in Hiroshima. Quantitative extraction of 100 to 400 μg nickel, containing only a few million atoms of ^{63}Ni , was required from gram-sized copper samples. Stable ^{63}Cu (69% natural abundance) is an isobar to ^{63}Ni and causes severe interference in the mass spectrometric measurements. For example, Hiroshima samples beyond 1,000 m from the hypocentre were expected to contain 18 orders of magnitude more ^{63}Cu than ^{63}Ni . Thus, the nickel had to be chemically separated from copper before AMS analysis could take place. Methods were developed consisting of electrochemical separation to remove the gross copper material followed by a procedure in which the nickel reacted with carbon monoxide to form nickel tetracarbonyl, a volatile compound which could be thermally decomposed to nickel metal directly in the AMS sample holder. Both of these processes are highly specific to nickel. Details of the methods are given elsewhere^{7,8}. Two laboratories were involved in the AMS measurements. At the Livermore facility, $^{63}\text{Ni}^{10+}$ ions were accelerated to 99 MeV and counted in a four-anode gas ionization detector; the overall system had a typical rejection factor of 5×10^5 for ^{63}Cu . At the Munich tandem laboratory, $^{63}\text{Ni}^{12+}$ ions were accelerated to 175 MeV. At this energy, a gas-filled magnet could be used to separate the isobars further. The use of a gas-filled magnet in combination with a position- and angle-sensitive multi-anode gas ionization detector resulted in a ^{63}Cu rejection factor of 10^9 . The achievement of this high rejection factor for ^{63}Cu allowed measurements of distant samples and also of close samples high in stable nickel. Both systems have been described in more detail previously^{8–11}.

Suitable copper samples had to have: (1) a known location at the time of the bombing, (2) been in line-of-sight of the explosion with no or minimal intervening shielding, (3) been located at distances most relevant to the atomic-bomb survivor health studies, and (4) be low in stable nickel to keep the competing production from the thermal $^{62}\text{Ni}(\text{n},\gamma)^{63}\text{Ni}$ reaction as low as possible and thereby maximize the $^{63}\text{Ni}/\text{Ni}$ ratio. Bomb-exposed copper samples have been collected from seven distances in Hiroshima ranging from 380 m to more than 5,000 m from the hypocentre. Each of these samples has been measured two or more times. The samples from 380 to 1,461 m provide direct information on the fast-neutron fluences while the samples from 1,880 and 5,062 m determine the

Table 1 ^{63}Ni activation in copper samples in Hiroshima

Sample location (distance from hypocentre)	Description of samples	Number of measurements	^{63}Ni per g Cu measured ($\times 10^4$)	^{63}Ni per g Cu measured-background† ($\times 10^4$)	^{63}Ni per g Cu calculated‡ ($\times 10^4$)
Bank of Japan (380 m)	Lightning conductor, unshielded	3	400 ± 40	580 ± 50	904
Soy Sauce Brewery (949 m)	Lightning conductor, inside an iron pipe (0.35 cm Fe shielding)	2	44 ± 14	54 ± 20	50
City Hall (1,014 m)	Lightning conductor, inside an iron pipe (0.40 cm Fe shielding)	6	26.5 ± 2.7	28 ± 5	30
Elementary school (1,301 m)	Rain gutter, unshielded	3	11.0 ± 1.4	$5.4 (+4.1, -3.4)$	5.6
Radioisotope building (1,461 m)	Rain gutter, unshielded	3	10.3 ± 1.7	$4.5 (+4.5, -3.8)$	3.0
Sumitomo Bank§ (1,880 m)	Rain gutter, unshielded	2	$7.3 (+2.6, -2.1)$		
Kusatsu Hachiman Shrine§ (5,062 m)	Copper roof, unshielded	2	$7 (+8, -5)$		

Results of measured and calculated (DS86) ^{63}Ni activation (given numbers are rounded). The hypocentre is the location on the ground directly under the bomb explosion. According to DS86, the Hiroshima bomb was exploded at 580 m above the hypocentre. Each sample was measured two or more times. The City Hall sample was measured using both the LLNL and the Munich AMS facility. The others were measured using the Munich facility only.

*Uncertainties correspond to 68% confidence intervals for Poisson-distributed data³⁰.

†A background of 7.3×10^4 ^{63}Ni per g Cu was subtracted from the measured values and the results corrected for decay since 1945.

‡ ^{63}Ni per g Cu calculated to have been produced in each sample by Hiroshima bomb neutrons based on the DS86 neutron output and energy spectrum, including both $^{63}\text{Cu}(\text{n,p})^{63}\text{Ni}$ and $^{62}\text{Ni}(\text{n},\gamma)^{63}\text{Ni}$ reactions.

§The weighted average of these two large-distance samples was used for background subtraction.

^{63}Ni content in copper samples not exposed significantly to bomb neutrons.

The calculations of the activation of ^{63}Ni in copper samples were based on the DS86⁶ neutron output (angle and spectrum) and the DS86 aboveground fluences. The *in situ* sample calculations were performed using a Monte Carlo adjoint shielding code²⁸. The calculations derived the number of ^{63}Ni atoms created in the copper sample by: (1) transporting local DS86 above ground fluences in air, sample, and surrounding materials taking into account scattering and absorption, and (2) taking into account the energy-dependence of the cross-sections²⁹ for the two relevant reactions, that is, $^{63}\text{Cu}(n,p)^{63}\text{Ni}$ (fast neutrons) and $^{62}\text{Ni}(n,\gamma)^{63}\text{Ni}$ (thermal neutrons). The fraction of ^{63}Ni produced by thermal neutrons in the copper samples was calculated to be 24% (380 m), 5.7% (949 m), 0.2% (1,014 m), 0.3% (1,301 m), and 0.1% (1,461 m), respectively.

The description of the Hiroshima copper samples and the measurement and calculation results are listed in Table 1. The measured results are expressed as mean ^{63}Ni atoms per g Cu values with uncertainties corresponding to 68% confidence intervals. The statistical uncertainty was obtained using a well-established procedure for Poisson-distributed data³⁰ (which was the case here for the AMS measurements) and was propagated from the uncertainties associated with the measured $^{63}\text{Ni}/\text{Ni}$ ratios, the measured mass of the Cu samples, the measured total Ni concentrations of the copper samples, and the measured background of ^{63}Ni atoms per g Cu in unexposed copper samples. As these are taken to be independent variables, the uncertainties were propagated in quadrature. The results show a consistent set of measurements with the ^{63}Ni atoms per g Cu values decreasing sharply with increasing distance from the bomb explosion and approaching a constant background beyond about 1,800 m (Table 1). The constant background beyond about 1,800 m appears to originate from a combination of *in situ* production of ^{63}Ni due to cosmic rays, sample processing, and AMS measurement. The next-to-last column lists the measured ^{63}Ni atoms per g Cu ratios after subtracting a background value of $7.3 (+2.4, -1.9) \times 10^4$ atoms of ^{63}Ni per g Cu and correction for radioactive decay since 1945. Correction for radioactive decay of ^{63}Ni was done to provide comparisons with calculations at the time of bombing. Listed in the last column are the results from the sample-specific DS86 calculations at the time of bombing.

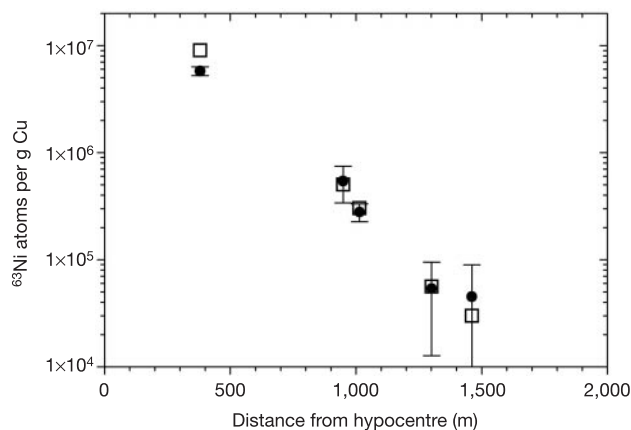


Figure 1 Measured and calculated ^{63}Ni produced in copper samples by the bomb neutrons in Hiroshima. The measured bomb-induced activation (solid circles) was obtained by first subtracting the background of 7.3×10^4 atoms ^{63}Ni per g Cu from the total measured values and then correcting for the radioactive decay of ^{63}Ni (half-life, 100.1 ± 2.0 yr) since 1945. The measurements are compared with results from sample-specific computer modelling calculations based on the DS86 dosimetry system (open squares).

These data and the results of the calculations are plotted in Fig. 1. They demonstrate that our ^{63}Ni measurements are in good agreement with those predicted from DS86 at distances between 900 and 1,500 m. This is particularly significant because that is the range of distances that dominate the atomic-bomb-survivor data and hence the radiation risk estimates for human cancer. The measurement at 380 m is somewhat below the DS86 calculation, suggesting that some downward revision in the DS86 neutron doses should be considered at distances closer to the hypocentre. The result for ^{63}Ni at 380 m has been compared with the results available for ^{32}P at similar distances. Six ^{32}P measurements at distances ranging from 331 to 433 m were obtained from an evaluation of the ^{32}P data in Hiroshima²³. Expressing these measurements as measured-to-calculated ratios, that is, (measured ^{32}P)/(calculated ^{32}P) based on DS86, results in a mean for the six ^{32}P measurements of 0.75 with a standard deviation of 0.12. This is compared with our ^{63}Ni measurement at 380 m, which has a measured-to-calculated ratio of 0.64 ± 0.06 . From this comparison it is observed that the fast neutron measurements at about 400 m from the hypocentre are in reasonable agreement with each other and both sets of measurements are lower than DS86. These results may be consistent with a slightly underestimated height-of-burst for the Hiroshima bomb.

In conclusion, good agreement of our experimental results for fast-neutron fluences with DS86 calculations is observed at the most important distances between 900 and 1,500 m. Given the close correlation between fast-neutron fluence and neutron dose, it is inferred from these results that neutron dose discrepancies, if there are any, should also be small at these distances. Taken together with previous findings that the gamma rays were adequately validated in Hiroshima and Nagasaki using thermoluminescence dosimetry (TLD)⁶ and that the neutrons appear to be consistent with DS86 in Nagasaki¹⁹, these results provide a strong measurement basis for the Hiroshima dosimetry and suggest that any future revisions in the doses for atomic-bomb survivors should be minor. □

Received 24 March; accepted 28 May 2003; doi:10.1038/nature01815.

1. BEIR V Report. *Health Effects of Exposure to Low Levels of Ionizing Radiation* (National Academy Press, Washington DC, 1990).
2. Hoshi, M. *et al.* Europium-152 activity induced by Hiroshima atomic bomb neutrons: comparison with the ^{32}P , ^{60}Co , and ^{152}Eu activities in Dosimetry System 1986 (DS86). *Health Phys.* **57**, 831–837 (1989).
3. Straume, T. *et al.* Neutron discrepancies in the DS86 Hiroshima dosimetry system. *Health Phys.* **63**, 421–426 (1992).
4. Shizuma, K. *et al.* Residual ^{152}Eu and ^{60}Co activities induced by neutrons from the Hiroshima atomic bomb. *Health Phys.* **65**, 272–282 (1993).
5. Shizuma, K. *et al.* Residual ^{60}Co activity in steel samples exposed to the Hiroshima atomic-bomb neutrons. *Health Phys.* **75**, 278–284 (1998).
6. Roesch, W. C. (ed.) *US-Japan Joint Reassessment of Atomic Bomb Radiation Dosimetry in Hiroshima and Nagasaki -Final Report* (Radiation Effects Research Foundation, Hiroshima, 1987).
7. Marchetti, A. A. *et al.* Ultra-separation of nickel from copper metal for the measurement of ^{63}Ni by AMS. *Nucl. Instrum. Meth. B* **123**, 230–234 (1997).
8. McAninch, J. E. *et al.* Measurement of ^{63}Ni and ^{59}Ni by accelerator mass spectrometry using characteristic projectile X-rays. *Nucl. Instrum. Meth. B* **123**, 137–142 (1997).
9. Knie, K., Faestermann, T. & Korschinek, G. AMS at the Munich gas-filled analyzing magnet system GAMS. *Nucl. Instrum. Meth. B* **123**, 128–131 (1997).
10. Rugel, G. *et al.* Accelerator mass spectrometry of ^{63}Ni using a gas-filled magnet at the Munich Tandem Laboratory. *Nucl. Instrum. Meth. B* **172**, 934–938 (2000).
11. Rühm, W. *et al.* Accelerator mass spectrometry of ^{63}Ni at the Munich tandem laboratory for estimating fast neutron fluences from the Hiroshima atomic bomb. *Health Phys.* **79**, 358–364 (2000).
12. Hashizume, T. *et al.* Estimation of the air dose from the atomic bombs in Hiroshima and Nagasaki. *Health Phys.* **13**, 149–161 (1967).
13. Nakanishi, T., Imura, T., Komura, K. & Sakanoue, M. ^{152}Eu in samples exposed to the nuclear explosions at Hiroshima and Nagasaki. *Nature* **302**, 132–134 (1983).
14. Habersack, G. *et al.* Accelerator mass spectrometry with fully stripped ^{36}Cl ions. *Radiocarbon* **28**, 204–210 (1986).
15. Korschinek, G. *et al.* Accelerator mass spectrometry with completely stripped ^{41}Ca and ^{53}Mn ions at the Munich tandem accelerator. *Nucl. Instrum. Meth. B* **29**, 67–71 (1987).
16. Straume, T. *et al.* Use of accelerator mass spectrometry in the dosimetry of Hiroshima neutrons. *Nucl. Instrum. Meth. B* **52**, 552–556 (1990).
17. Kato, K. *et al.* Accelerator mass spectrometry of ^{36}Cl produced by neutrons from the Hiroshima bomb. *Int. J. Radiat. Biol.* **58**, 661–672 (1990).
18. Rühm, W., Kato, K., Korschinek, G., Morinaga, H. & Nolte, E. ^{36}Cl and ^{41}Ca depth profiles in a Hiroshima granite stone and the dosimetry system 1986. *Z. Phys. A* **341**, 235–238 (1992).
19. Straume, T., Harris, L. J., Marchetti, A. A. & Egbert, S. D. Neutrons confirmed in Nagasaki and at the Army Pulsed Radiation Facility: implications for Hiroshima. *Radiat. Res.* **138**, 193–200 (1994).

20. Rühm, W. *et al.* The dosimetry system DS86 and the neutron discrepancy in Hiroshima—historical review, present status, and future options. *Radiat. Environ. Biophys.* **37**, 293–310 (1998).
21. Arakatsu, B. *Collection of Investigative Reports on Atomic Bomb Disaster* 34–35 (Science Council of Japan, Tokyo, 1953).
22. Yamasaki, F. & Sugimoto, A. *Collection of Investigative Reports on Atomic Bomb Disaster* 18–19 (Science Council of Japan, Tokyo, 1953).
23. Gritzner, M. L. & Woolson, W. A. in *US-Japan Joint Reassessment of Atomic Bomb Radiation Dosimetry in Hiroshima And Nagasaki-Final Report* Vol. 2 (ed. Roesch, W. C.) 283–292 (Radiation Effects Research Foundation, Hiroshima, 1987).
24. Barnes, I. L., Garfinkel, S. B. & Mann, W. B. Standardization of ^{63}Ni by efficiency tracing. *Int. J. Appl. Radiat. Isot.* **22**, 781–783 (1971).
25. Marchetti, A. A. & Straume, T. A search for neutron reactions that may be useful for Hiroshima dose reconstruction. *Appl. Radiat. Isotop.* **47**, 97–103 (1996).
26. Straume, T., Marchetti, A. A. & McAninch, J. E. New analytical capability may provide solution to the neutron dosimetry problem in Hiroshima. *Radiat. Prot. Dosim.* **67**, 5–8 (1996).
27. Shibata, T. *et al.* A method to estimate the fast-neutron fluence for the Hiroshima atomic bomb. *J. Phys. Soc. Jpn* **63**, 3546–3547 (1994).
28. Johnson, J. O. (ed.) *A User's Manual for MASH v2.0—Monte Carlo Adjoint Shielding Code System* Report ORNL/TM/11778/R2 (Oak Ridge National Laboratory, Oak Ridge, TN, 1999).
29. Rose, P. F. (ed.) *ENDF/B-VI Summary Documentation, Cross-Section Evaluation Group Report BNL-NCS-17541* (Nuclear Data Center, Brookhaven National Laboratory, Upton, NY, 1991).
30. Feldman, G. J. & Cousins, R. D. Unified approach to the classical statistical analysis of small signals. *Phys. Rev. D* **57**, 3873–3889 (1998).

Acknowledgements We thank I. Proctor and A. M. Kellerer for their essential support of this project. We also thank S. Fujita and K. Shizuma for providing copper samples for this study. We thank the following organizations for supporting this work: the US Department of Energy, the US National Academy of Sciences, the European Commission, the German Federal Ministry of Environment, Nature Conservation and Nuclear Safety.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.S. (Straume1@aol.com).

Niche lability in the evolution of a Caribbean lizard community

Jonathan B. Losos¹, Manuel Leal^{2*}, Richard E. Glor¹, Kevin de Queiroz³, Paul E. Hertz⁴, Lourdes Rodríguez Schettino⁵, Ada Chamizo Lara⁵, Todd R. Jackman⁶ & Allan Larson¹

¹Department of Biology, Campus Box 1137, Washington University, St. Louis, Missouri 63130, USA

²Department of Biological Sciences, Union College, Schenectady, New York 12308, USA

³Division of Amphibians and Reptiles, National Museum of Natural History, Smithsonian Institution, Washington, DC 20560, USA

⁴Department of Biology, Barnard College, 3009 Broadway, New York 10027, USA

⁵Instituto de Ecología y Sistemática, CITGMA, Carretera de Varona km 3.5, Boyeros, La Habana 10800, Apartado Postal 8029, Cuba

⁶Department of Biology, Villanova University, Villanova, Pennsylvania 19085, USA

* Present address: Department of Biological Sciences, Vanderbilt University, VU Station B 351634 Nashville Tennessee 37235, USA

Niche conservatism—the tendency for closely related species to be ecologically similar—is widespread^{1–3}. However, most studies compare closely related taxa that occur in allopatry³; in sympatry, the stabilizing forces that promote niche conservatism^{4,5}, and thus inhibit niche shifts, may be countered by natural selection favouring ecological divergence to minimize the intensity of interspecific interactions^{6,7}. Consequently, the relative importance of niche conservatism versus niche divergence in determining community structure has received little attention⁷. Here, we examine a tropical lizard community in which species have a long evolutionary history of ecological interaction. We find that evolutionary divergence overcomes niche conservatism: closely related species are no more ecologically similar than expected by

random divergence and some distantly related species are ecologically similar, leading to a community in which the relationship between ecological similarity and phylogenetic relatedness is very weak. Despite this lack of niche conservatism, the ecological structuring of the community has a phylogenetic component: niche complementarity only occurs among distantly related species, which suggests that the strength of ecological interactions among species may be related to phylogeny, but it is not necessarily the most closely related species that interact most strongly.

Anolis lizards are a dominant component of Caribbean ecosystems (reviewed in refs 8 and 9) and are well suited for studies of the evolution of community structure because the species on individual islands have a long history of interaction and coevolution. For example, 55 of 58 species on Cuba are endemic (the remaining three have colonized other Caribbean islands from Cuba), and most are members of large clades that have diversified on Cuba¹⁰. Species on many islands attain extremely high densities^{11,12}, and many species—differing in ecology, morphology, and behaviour—coexist locally⁸. Interactions among sympatric species can be strong^{8,9,13,14}, usually as a result of interspecific competition, although intra-guild predation may sometimes be important¹⁵.

We studied the community structure of anoles at Soroa, Biosphere Preserve Sierra del Rosario, in the Pinar del Río province of western Cuba. Eleven anole species occur sympatrically at Soroa, the highest anole diversity known from any island or continental site. Of these species, ten are either widely distributed in Cuba or are members of island-wide clades of ecologically similar species (for example, the *Anolis equestris* group, to which *A. luteogularis* belongs, occurs throughout Cuba and is composed of six primarily allopatric species similar in morphology and ecology). Because the clades of Cuban anoles to which the Soroa species belong are widespread and arose within a relatively short period in the distant past¹⁰ (Fig. 1), the sympatric clades at Soroa have probably coexisted for a long time and over a large spatial scale. Thus, these *Anolis* species probably evolved in the presence of the same clades with which they currently coexist, a necessary prerequisite for community coevolution.

We examined ecological relationships among these species to investigate whether the community exhibited nonrandom ecological or phylogenetic structure. We measured ecological variables relevant to the three resource axes that sympatric *Anolis* generally partition: structural habitat, thermal habitat, and prey size¹⁶. Principal components analysis reveals three significant axes of ecological differentiation (Table 1; results below are qualitatively unchanged if another, nearly significant, axis is also retained). Examination of the position of species in multivariate ecological space reveals both that niche use has not been conserved and that the community is nonrandomly structured (Fig. 2).

The minimal extent of niche conservatism is indicated by the weak association between phylogenetic relationship and position in multivariate ecological space: phylogenetic similarity explains less than 4% of the variation in ecological similarity among species (Mantel test, $P = 0.11$ – 0.30 depending on phylogenetic topology and mode of character evolution used in the analysis; all variables but one exhibit similarly low correlations with phylogenetic relationships (Table 1); P -values in Mantel tests based on 5,000 simulations). The molecular data strongly reject alternative phylogenetic topologies in which ecologically similar species are grouped phylogenetically (see Supplementary Information).

Although some closely related species differ little ecologically, many distantly related species are just as ecologically similar, and some closely related species are ecologically dissimilar (Fig. 2). Moreover, although members of the *sagrei* and *porcatus* clades form clusters in ecological space (Fig. 2; multivariate analysis of variance, MANOVA, Wilks' $\lambda = 0.013$, $F_{12,10} = 3.79$, $P = 0.018$), they are no more ecologically similar than would be expected for

Table 1 Principal components analysis

Component loadings	Principal component axes				Variation explained by phylogenetic relationships*
	1	2	3	4	
Activity time	0.793	-0.111	0.286	0.256	0-3%†
Body temperature	0.688	0.151	0.400	0.455	0-12%†
Perch height	0.791	0.396	-0.294	-0.275	3-5%
Perch diameter	-0.204	0.900	0.298	-0.152	0-3%†
Use of rocks	-0.852	-0.069	0.049	0.455	2-5%
Snout-vent length	0.141	0.514	-0.715	0.442	12-42%
Diameter of surfaces used during locomotion	-0.310	0.893	0.214	0.012	3-10%†
Variance explained by components	1	2	3	4	
	2.612	2.069	0.977	0.773	
Percentage of total variance explained‡	1	2	3	4	
	37.3	29.6	14.0	11.0	
	(37.0)	(22.8)	(15.6)	(10.9)	

Activity time was the weighted average of eight evenly spaced categories (given values of 1-8) corresponding to the times of activity transects (for example, category 1 = 07:00-08:30, category 2 = 08:30-10:00). For perch height, perch diameter, diameter of substrates used during locomotion, and body temperature, mean values were used for each species. These variables, as well as snout-vent length, were natural-log-transformed before analysis. 'Use of rocks' was the proportion of all animals observed on or within 1 m of a rocky surface and was arcsine square-root-transformed. Snout-vent length values for adult males were taken from ref. 27; when ranges were provided, the value closest to our measurements of specimens from Soroa was used.

*Proportion of similarity among species in these traits explained by phylogenetic similarity. Ranges indicate results using different phylogenetic topologies and branch lengths.

†Negative correlations between trait and phylogenetic similarity.

‡Expected by chance according to Broken Stick model in parentheses; thus, axes 1, 2 and 4 are statistically significant.

clades of their age (phylogenetic simulations: *sagrei* clade, $P = 0.49-0.76$ (ranges represent results from analyses using different phylogenies, branch lengths and modes of character evolution); *porcatus* clade, $P = 0.60-0.80$; average for both clades, $P = 0.52-0.73$). Taken together, these results indicate that recently diverged species are ecologically similar, but their divergence is not constrained and thus niches are conserved no more than would be expected if they were diverging randomly. Moreover, the data indicate that ecologi-

cal divergence is occurring at a sufficiently rapid rate that only the ecological similarities among relatively closely related species have a significant component attributable to common ancestry; overall comparisons that include distantly related species reveal no correlation between phylogenetic relatedness and ecological similarity.

Nonrandom community structure is evident in the pattern of similarity of species along different ecological axes. The niche complementarity hypothesis¹⁶ predicts that species similar along one niche axis will differ along others. The positions of species along the three significant principal components (PC) axes indicate that niche complementarity occurs at Soroa (Table 2). For all three pairwise combinations of the three axes, the distance separating species in ecological space along one axis is negatively correlated with the corresponding distance along the second axis. The probability that all three pairwise correlations should have as strong a negative relationship as that observed is extremely low ($P \leq 0.002$; Table 2).

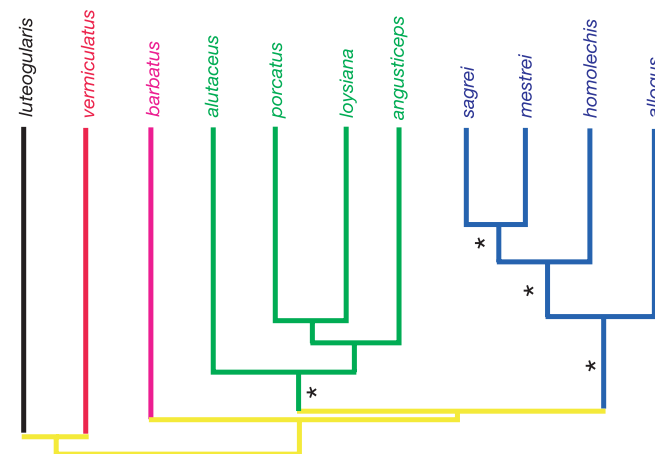


Figure 1 Phylogenetic relationships of the 11 anole species at Soroa. The phylogeny is a subset of a phylogeny for 129 species of anoles with non-Soroa species removed. Branch lengths are proportional to time since divergence (illustrated here is the maximum-likelihood tree with branch lengths fitted by non-parametric rate smoothing). Asterisks indicate clades that are strongly supported on both the maximum likelihood (>90% bootstrap support) and bayesian (100%) trees in analyses including the 11 Soroa species. The same nodes are recovered on the maximum parsimony tree, all but the *porcatus* clade (green) with high (>94% bootstrap) support. In addition, both the *porcatus* and the *sagrei* (blue) clades were originally described on the basis of morphological data. Maximum Tamura-Nei corrected distance observed between species is 0.31, which corresponds to a divergence time of approximately 24 Myr ago²⁸. This is in accordance with dates for anole divergence based on immunological differences²⁹ and fossil amber specimens³⁰. Of the Soroa species, all are endemic to Cuba except *A. sagrei*, which has colonized the Bahamas and a number of other islands in the western Caribbean, and all are either widespread or members of widespread clades of ecologically similar species except *A. vermiculatus*, which is only found in western Cuba.

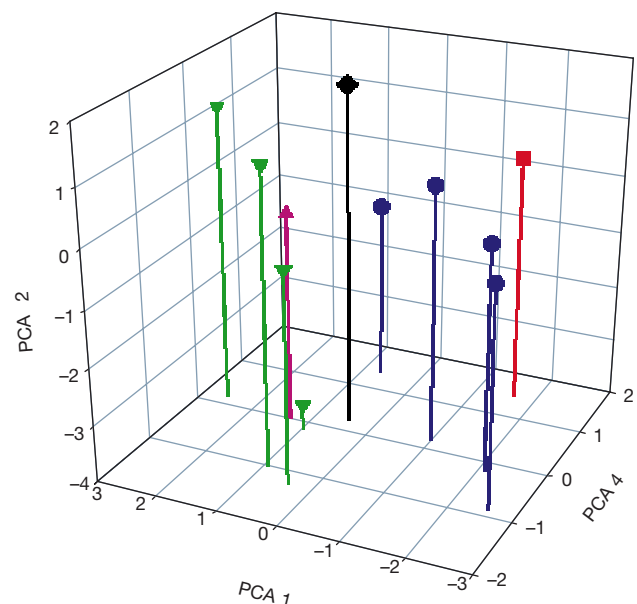


Figure 2 Dispersion of species in multivariate ecological space. Colours and symbols correspond to the five clades in Fig. 1. PCA, principal components analysis.

Table 2 **Niche complementarity**

	PC 1 versus 2	PC 1 versus 4	PC 2 versus 4	Probability*
All species	-0.19	-0.18	-0.23	0–2/1,000
Within <i>sagrei</i> clade	0.40	0.90	0.34	1,000/1,000†
Between <i>sagrei</i> clade species and other species	-0.21	-0.26	-0.24	1–16/1,000
Within <i>porcatus</i> clade	-0.41	0.38	-0.37	175–191/1,000
Between <i>porcatus</i> clade species and other species	-0.11	-0.23	-0.22	3–16/1,000

*Statistical significance was judged using the number of simulations out of 1,000 in which the number of negative pairwise correlations with an absolute value greater than that for the smallest negative pairwise correlation in the real data was as great as the number of negative pairwise correlations in the real data. Numbers indicate the range of results using the different phylogenies and branch-length estimations. Thus, for 'all species', in no more than 2/1,000 simulations did all three pairwise comparisons have correlations less than -0.18, whereas for the *porcatus* clade, 17.5–19.1% of simulations produced at least two correlations less than -0.37.

†Given that all pairwise correlations are positive for the within-*sagrei* clade analysis, no evidence exists for niche complementarity. Indeed, in only 4.5–7.3% of simulations were all three correlations positive with a correlation greater than 0.34.

In all cases, the results are the same if PC 3, which does not explain a statistically significant proportion of the variance (Table 1), is included. For example, for those cases in which all three comparisons are in the same direction (that is, have the same sign), the three comparisons involving PC 3 are also in that direction and the statistical significance is qualitatively unchanged. For the comparison of *porcatus* group species to other species, two of three comparisons involving PC 3 have a negative sign.

Consideration of niche dissimilarity in a phylogenetic context provides insight to the evolution of community structure. Niche complementarity is not evident within either the *porcatus* or *sagrei* clades; indeed, in the *sagrei* clade, all pairwise correlations are positive, implying that degree of ecological differentiation among species is positively related among ecological axes (Table 2). By contrast, examination of distances separating members of each clade from other members of the community reveals significant niche complementarity for both clades (Table 2). Thus, although community structure shows little evidence of phylogenetic effects among distantly related species, interactions among these species appear to be responsible for the nonrandom dispersion of species in ecological space.

These results have important implications for studies of community ecology. The increasing availability of detailed phylogenetic information at last permits explicit examination of the role of evolutionary divergence in community assembly⁷. In the case of the anoles of Soroa, and in contrast to the many studies that have documented niche conservatism³, the limited extent of phylogenetic structuring in this community implies that niche divergence, rather than conservatism, has had primacy in shaping community evolution. Although the community is not structured phylogenetically, it is structured ecologically: the nonrandom distribution of species in multivariate ecological space suggests that interspecific interactions have shaped community structure, a hypothesis that is supported by the extensive evidence that indicates that ecologically similar anole species interact strongly^{8,9,13,14}.

This ecological structuring, in turn, reveals an unexpected phylogenetic aspect. Since Darwin's time, ecologists have suggested that interspecific interactions may be strongest among closely related species^{17,18}. However, when such strong interspecific interactions occur in a system dominated by niche divergence, species may diverge from near relatives to the extent that they interact just as strongly with less related species. Consequently, the evolutionary outcome may be that ecological interactions among distantly related species play an important role in structuring communities⁶, a prediction which is supported by our discovery that the non-random structure of the Soroa community results from the ecological spacing of distant relatives.

Our findings are particularly relevant to the study of adaptive radiation, a topic of considerable recent interest⁶. A classic model of adaptive radiation postulates that interspecific interactions play an important role in driving evolutionary divergence⁶. Although adaptive radiations have been studied from both community ecological and evolutionary perspectives, integration of these approaches is just beginning⁶. Consequently, investigation of the generality of the patterns exhibited by anoles will provide insight into how adaptive radiation leads to the evolution of diverse communities.

Quantitative analyses of the relationship between ecological and phylogenetic similarity do not yet exist for other evolutionarily

assembled communities, but qualitative examination suggests both that substantial variation may exist among adaptive radiations and that the evolutionary age of the radiation may be related to this variation. For example, relatively young evolutionary radiations of Darwin's finches¹⁹ and Lake Malawi cichlids²⁰ have produced communities in which closely related species are ecologically similar; by contrast, in a community of Old World leaf warblers, which like anoles are the result of a more ancient (>10 Myr ago) radiation, ecological and phylogenetic similarity do not appear to be related²¹, as is also the case in anoles. These differences suggest that community evolution may occur in fundamentally different ways; the length of time community members have been coevolving may be important, but other factors such as prevalence of sexual selection or geographic setting also should be investigated. Future study of ecological relationships in a phylogenetic context will permit testing of these hypotheses and will enrich our understanding of both adaptive radiation and the evolutionary genesis of community structure. □

Methods

Ecological data collection

Habitat-use data were collected by walking transects in the forest in May 1997 and noting perch height, type, and diameter for each undisturbed adult lizard observed. Data were also collected for *A. sagrei*, an edge- and open-habitat species, in areas surrounding the forest. Diameter of substrates used during locomotion was recorded by observing adult lizards from a distance >5 m for 5–257 min and noting the diameter of every woody surface used; multiple measurements were taken when lizards moved along surfaces that varied greatly in diameter. Observations (made 07:00–18:00 when rain was not falling or imminent) were conducted only on males for species in which sex could be determined from a distance. For each individual, we calculated the mean of all substrate diameters.

Body temperature and activity time were measured in a pre-arranged regimen in which sampling effort was constant throughout the day. Lizards were captured and cloacal temperature quickly recorded by inserting a thermocouple using standard methods. Activity time was measured by walking transects at 1.5 h intervals throughout two days and noting every lizard observed. For rare species, data were augmented by measurements taken whenever individuals were observed. Because of Soroa's protected status, we did not collect the large sample of specimens necessary for diet analyses. Instead, we used snout-vent length, which correlates strongly with prey size in *Anolis*⁸. Except where noted, data were collected for both males and females. Although anoles are sexually dimorphic in habitat use²², dimorphism in Soroa species was minor relative to interspecific differences.

Phylogenetic analysis

We used PAUP* v4.0b10 (ref. 23) and MrBayes v3.0b3 (ref. 24) to generate phylogenies from mitochondrial DNA haplotypes representing 129 anole species using parsimony, maximum-likelihood and bayesian approaches. DNA extraction, amplification and sequence alignment were conducted as described in ref. 10, which presented 53 of these sequences. We sequenced approximately 1,500 base pairs, including ND2, five transfer RNAs, the origin of light-strand replication, and part of CO1. The HKY model (transition/transversion ratio = 3, proportion of invariable sites = 0.2, shape parameter = 0.7), selected for maximum-likelihood analyses using Modeltest 3.0 (ref. 25), resulted in a single tree with a likelihood score of $-\ln(66946.90)$. For the bayesian analysis, four chains were run for 1,000,000 generations. Following a burn-in period of 50,000 generations, the mean likelihood score for sampled trees was $-\ln(66917.60)$ (s.d. = 11.99). Relative dates of divergence of the species present at Soroa were estimated on both trees by transforming branch lengths with the Langley-Fitch method²⁶ and non-parametric rate smoothing using the program 'r8s' version 1.01b (<http://ginger.ucdavis.edu/r8s/>).

Statistical analysis

All analyses were conducted on four ultrametric trees (Langley–Fitch and NPRS branch lengths for both the maximum-likelihood and bayesian phylogenies); further analyses were run assuming both a speciation and a gradual mode of character evolution. Results from these analyses were qualitatively nearly identical.

To examine whether distances between species in ecological space was related to phylogenetic proximity, we conducted a Mantel test comparing matrices of ecological and phylogenetic distance. Ecological distance was the euclidean distance between species in a multidimensional space determined by the axes of a PC analysis. Phylogenetic distances were the branch lengths separating species (that is, their patristic distance). To examine whether niche position of the species in the two species-rich clades (Fig. 1) was conserved, we examined whether members of these clades occupied a nonrandomly small part of ecological space relative to the other species by conducting a MANOVA with PC axis scores as variables and five groups (the two clades and the other three species).

Then, to assess whether members of these clades were more similar to each other than would be expected by random evolutionary divergence, we calculated for each clade the ratio of the mean distance between species within the clade and the mean distance of species within the clade to other species. To assess whether these ratios were unusually small, we conducted phylogenetic simulations. For each significant PC axis, trait evolution was simulated on the phylogeny assuming either gradual or speciation models of character evolution; variance in trait value among species in the simulations was adjusted to that observed in the real data. Using these simulations, we then assessed whether the ratios for each clade separately or for the average of the two clades was significantly smaller than expected by chance.

To determine whether species were nonrandomly dispersed in ecological space, we examined whether the pairwise euclidean distance between species on one PC axis was negatively related to the distance along a second axis ('niche complementarity'). Because the three pairwise comparisons among the three PC axes are not independent, an experiment-wise *P*-value cannot be calculated. We investigated the statistical significance of these findings using phylogenetic simulation as above. For both the real and simulated values, correlations were calculated between euclidean distances on each possible pairwise combination of PC axis scores.

Received 29 January; accepted 29 May 2003; doi:10.1038/nature01814.

- Peterson, A. T., Soberón, J. & Sánchez-Cordero, V. Conservatism of ecological niches in evolutionary time. *Science* **285**, 1265–1267 (1999).
- Prinzing, A., Durka, S., Klotz, W. & Brandl, R. The niche of higher plants: evidence for phylogenetic conservatism. *Proc. R. Soc. Lond. B* **268**, 2383–2389 (2001).
- Webb, C. O., Ackerly, D. D., McPeck, M. A. & Donoghue, M. J. Phylogenies and community ecology. *Annu. Rev. Ecol. Syst.* **33**, 475–505 (2002).
- Harvey, P. H. & Pagel, M. D. *The Comparative Method in Evolutionary Biology* (Oxford Univ. Press, Oxford, 1991).
- MacArthur, R. H. *Geographical Ecology* (Princeton Univ. Press, Princeton, 1972).
- Schluter, D. *The Ecology of Adaptive Radiation*. (Oxford Univ. Press, Oxford, 2000).
- Silvertown, J., Dodd, M. & Gowing, D. Phylogeny and the niche structure of meadow plant communities. *J. Ecol.* **89**, 428–435 (2001).
- Losos, J. B. Integrative approaches to evolutionary ecology: *Anolis* lizards as model systems. *Annu. Rev. Ecol. Syst.* **25**, 467–493 (1994).
- Roughgarden, J. *Anolis Lizards of the Caribbean: Ecology, Evolution, and Plate Tectonics* (Oxford Univ. Press, 1995).
- Jackman, T. R., Larson, A., de Queiroz, K. & Losos, J. B. Phylogenetic relationships and tempo of early diversification in *Anolis* lizards. *Syst. Biol.* **48**, 254–285 (1999).
- Schoener, T. W. & Schoener, A. Densities, sex ratios, and population structure in four species of Bahamian *Anolis* lizards. *J. Anim. Ecol.* **49**, 19–53 (1980).
- Reagan, D. P. Congeneric species distribution and abundance in a three-dimensional habitat: the rain forest anoles of Puerto Rico. *Copeia*, **1992**, 392–403 (1992).
- Leal, M., Rodríguez-Robles, J. A. & Losos, J. B. An experimental study of interspecific interactions between two Puerto Rican *Anolis* lizards. *Oecologia* **117**, 273–278 (1998).
- Losos, J. B. & Spiller, D. Differential colonization success and asymmetrical interactions between two lizard species. *Ecology* **80**, 252–258 (1999).
- Gerber, G. P. & Echternacht, A. C. Evidence for asymmetrical intraguild predation between native and introduced *Anolis* lizards. *Oecologia* **124**, 599–607 (2000).
- Schoener, T. W. Resource partitioning in ecological communities. *Science* **185**, 27–39 (1974).
- Darwin, C. *On the Origin of Species* Ch. 3 (Murray, London, 1859).
- Hutchinson, G. E. *The Ecological Theater and the Evolutionary Play* Ch. 2 (Yale Univ. Press, New Haven, 1965).
- Petren, K., Grant, B. R. & Grant, P. R. A phylogeny of Darwin's finches based on microsatellite DNA length variation. *Proc. R. Soc. Lond. B* **266**, 321–329 (1999).
- Danley, P. D. & Kocher, T. D. Speciation in rapidly diverging systems: lessons from Lake Malawi. *Mol. Ecol.* **10**, 1075–1086 (2001).
- Richman, A. D. & Price, T. Evolution of ecological differences in the Old World leaf warblers. *Nature* **355**, 817–821 (1992).
- Butler, M. A. & Losos, J. B. Multivariate sexual dimorphism, sexual selection, and adaptation in Greater Antillean *Anolis* lizards. *Ecol. Monogr.* **72**, 541–559 (2002).
- Swofford, D. *PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods)* (Sinauer, Sunderland, MA, 2002).
- Huelsenbeck, J. P. & Ronquist, F. MRBAYES: Bayesian inference of phylogeny. *Bioinformatics* **17**, 754–755 (2001).
- Posada, D. & Crandall, K. A. Modeltest: testing the model of DNA substitution. *Bioinformatics* **14**, 817–818 (1998).
- Langley, C. H. & Fitch, W. An estimation of the constancy of the rate of molecular evolution. *J. Mol. Evol.* **3**, 161–177 (1974).
- Rodríguez Schettino, L. *The Iguanid Lizards of Cuba* (Univ. Press Florida, Gainesville, 1999).
- Macey, J. R. *et al.* Phylogenetic relationships among agamid lizards of the *Laudakia caucasia* species

- group: testing hypotheses of biogeographic fragmentation and an area cladogram for the Iranian Plateau. *Mol. Phylogenet. Evol.* **10**, 118–131 (1998).
29. Shochat, D. & Dessauer, H. C. Comparative immunological study of albumins of *Anolis* lizards of the Caribbean islands. *Comp. Biochem. Physiol.* **68A**, 67–73 (1981).
30. de Queiroz, K., Chu, L.-R. & Losos, J. B. A second *Anolis* lizard in Dominican amber and the systematics and ecological morphology of Dominican amber anoles. *Am. Mus. Nat. Hist. Novitates* **3249**, 1–23 (1998).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank J. Chase, R. Ricklefs and T. Schoener for comments, and V. Rivalta González, A. Daniel Alvarez, and A. Torres Barboza for assistance in the field. We thank the National Geographic Society, the David and Lucile Packard Foundation, the National Science Foundation and Barnard College for support.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.B.L. (losos@biology2.wustl.edu).

Constrained circulation at Endeavour ridge facilitates colonization by vent larvae

Richard E. Thomson¹, Steven F. Mihalý¹, Alexander B. Rabinovich^{1,2}, Russell E. McDuff³, Scott R. Veirs³ & Frederick R. Stahr³

¹Institute of Ocean Sciences, 9860 West Saanich Road, Sidney, British Columbia, V8L 4B2, Canada

²P.P. Shirshov Institute of Oceanology, Moscow 117851, Russia

³School of Oceanography, University of Washington, Seattle, Washington 98195-7940, USA

Understanding how larvae from extant hydrothermal vent fields colonize neighbouring regions of the mid-ocean ridge system remains a major challenge in oceanic research^{1,2}. Among the factors considered important in the recruitment of deep-sea larvae are metabolic lifespan, the connectivity of the seafloor topography, and the characteristics of the currents³. Here we use current velocity measurements from Endeavour ridge to examine the role of topographically constrained circulation on larval transport along-ridge. We show that the dominant tidal and wind-generated currents in the region are strongly attenuated within the rift valley that splits the ridge crest, and that hydrothermal plumes rising from vent fields in the valley drive a steady near-bottom inflow within the valley. Extrapolation of these findings suggests that the suppression of oscillatory currents within rift valleys of mid-ocean ridges shields larvae from cross-axis dispersal into the inhospitable deep ocean. This effect, augmented by plume-driven circulation within rift valleys having active hydrothermal venting, helps retain larvae near their source. Larvae are then exported preferentially down-ridge during regional flow events that intermittently over-ride the currents within the valley.

The Endeavour segment of Juan de Fuca ridge (Endeavour ridge) is a hydrothermally active, intermediate-rate spreading centre located in 2,500 m of water roughly 300 km seaward of British Columbia and Washington State in the northeast Pacific (Fig. 1). Hydrothermal venting from this segment is concentrated within a 1-km-wide, 10-km-long rift valley located along the ridge crest. Water depths within the axial valley shoal from 2,300 m in the south to 2,170 m in the north. Valley relief ranges from 100 to 150 m. A 25-km-wide, 50-km-long depression links the southern end of the Endeavour segment to the main portion of Juan de Fuca ridge.

(Topographic maps for Juan de Fuca ridge are available at http://ocean-ridge.ldeo.columbia.edu/ne_pac/imagery/imag_gif/nep112m.gif) The five major vent fields at Endeavour ridge are spaced 2–3 km apart along the axial rift valley⁴, with the apparently younger, highly active fields (Mothra, Main Endeavour, High Rise) located in the southern sector, and the older, weakly active fields (Salty Dawg, Sasquatch) located in the northern sector. The Main Endeavour field alone generates a total heat flux of 650 ± 100 megawatts (MW) through a combination of focused, high-temperature (350°C) venting and diffuse, low-temperature ($10\text{--}25^\circ\text{C}$) venting⁵. Vent plumes—as defined by their thermal, particulate and chemical anomalies—rise to a level of neutral buoyancy 50–350 m above bottom. Deeper portions of the plumes can remain trapped within the valley whereas those rising above the ridge crests are free to drift with the ambient flow.

Currents at Endeavour ridge and over the deep off-ridge site to the north (Fig. 1) consist of zero-mean oscillatory currents with amplitudes of about 10 cm s^{-1} superimposed on a near-steady background flow of about 5 cm s^{-1} . Oscillatory currents are in five primary frequency bands: (1) The semidiurnal band, principally the M_2 and S_2 tidal constituents with periods of 12.4 and 12.0 h, respectively; (2) the diurnal band, principally the K_1 and O_1 tidal constituents with periods of 23.9 and 25.8 h, respectively; (3) the wind-forced inertial (f) band with periods of approximately 16 h (frequencies a few per cent above the local Coriolis frequency, $f \approx 1.08 \times 10^{-4}\text{ s}^{-1}$); (4) the low-frequency ‘weather’ band spanning periods of days to weeks, with a broad spectral peak in the range of 4 to 6 days⁶; and (5) the high-frequency band arising mainly from nonlinear interactions among inertial and tidal frequencies⁷. Tidal currents account for roughly 50% of the flow variance.

The predominant M_2 tidal currents are aligned along-ridge (direction of strike $\approx 20^\circ$ true) with major tidal ellipse axes diminishing from about 5 cm s^{-1} above the ridge crest to $2\text{--}3\text{ cm s}^{-1}$ within the confines of the valley (Fig. 2). The predominant K_1 diurnal tidal currents are oriented cross-axis above the ridge, become increasingly clockwise rotary with proximity to the ridge crest, and then aligned along-axis within the axial valley. Diurnal currents are amplified by a factor of 3–4 at the ridge crest but are attenuated to 1 cm s^{-1} within the valley. Intermittent (roughly 1-week duration), wind-forced, inertial currents—often the strongest type of flow in the deep ocean—attain speeds of about 5 cm s^{-1} immediately above the ridge crest but are annihilated within the confines of the valley. Similarly, low-frequency currents within the wind-forced synoptic band are amplified by as much as a factor of seven near the ridge crest⁸, but are strongly damped within the valley.

Current meter data from 2000 and 2001, along with current records collected in the 1980s and 1990s, indicate a two-tier mean flow structure at Endeavour ridge (Fig. 3). Immediately above the ridge crest (Fig. 3a, top panel), and at elevations exceeding approximately 75–100 m above the valley floor (Fig. 3b, c), the slowly varying background flow is roughly along-axis towards the southwest and often exceeds 5 cm s^{-1} . However, periods of cross-axis velocity are common and often linked to eddy-like circulation (Fig. 3b). As one descends into the valley, the speed of the along-ridge current decreases and the flow becomes persistently into the valley. This ‘inflow’ draws cold bottom water towards the central buoyancy sources and is strongest (about 5 cm s^{-1}) in the southern and central sectors of the valley (Fig. 3c), where the valley is deepest and hydrothermal outputs are the highest. Inflow is weakest (about 1 cm s^{-1}) at the northern end, where the valley is shallowest and hydrothermal output is lowest. The shallow topographic saddle at the northern end of the valley, combined with the much higher thermal outputs from the southern and central vent fields, may explain why northward inflow extends well into the central valley from the south. Although the inflow is remarkably steady, it stalls or reverses direction (roughly 10% of the time at the central mooring

site) during intermittent bursts of strong southwesterly flow in the overlying water column. In autumn and winter, reversals occur every few weeks (Fig. 3a, lower panel) and persist for one or two days with along-axis flow speeds of 5 to 10 cm s^{-1} .

We can generalize our observations to other sectors of the global mid-ocean ridge system. Specifically, the ubiquitous diurnal tidal currents and wind-generated inertial currents of the deep ocean become amplified and more circularly polarized above the crests and flanks of oceanic ridges. If we view these amplified currents as turbulent-like fluctuations superimposed on a slowly varying mean flow, it follows that there is pronounced cross-axis dispersion of passively drifting larvae above oceanic ridge crests. (Here, dispersion arises from random velocity fluctuations, as well as from heterogeneity in the tidal and inertial flow fields.) Even if these topographically enhanced current speeds diminish slowly as L^{-1}

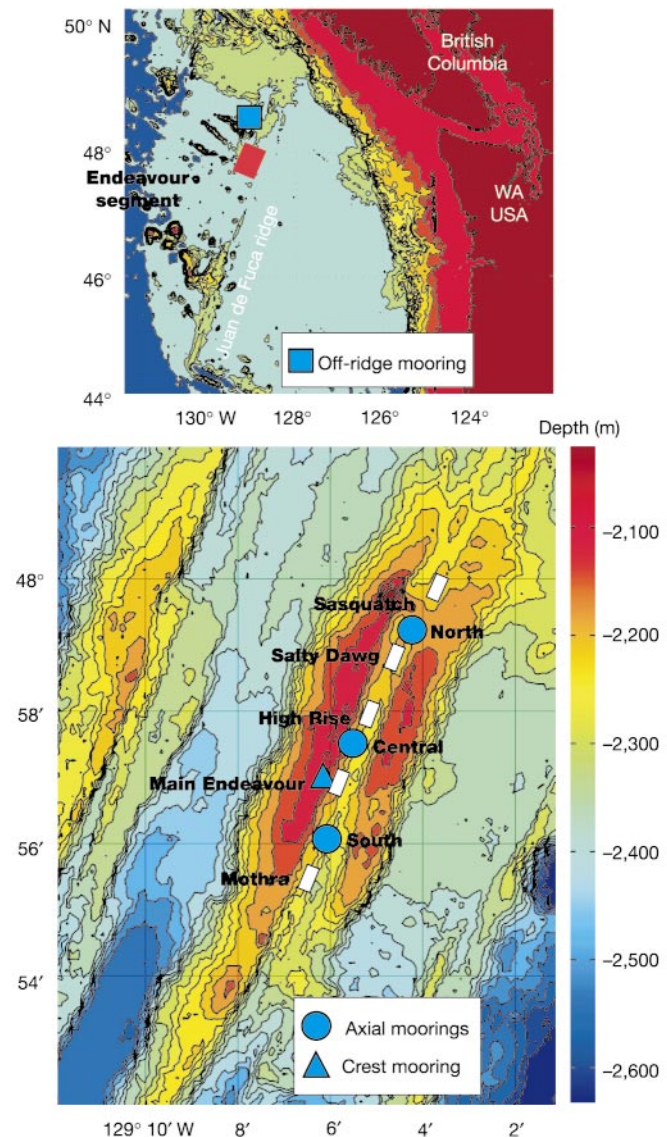


Figure 1 Chart of Endeavour ridge with an expanded view of the axial valley region. Small rectangles denote the five major vent sites within the valley. Circles denote the valley sites for the south mooring in the summer of 2000 and north, central and south moorings in the summer of 2001. The triangle denotes the 1996–97 bottom mooring site on the western ridge crest; the square in the top panel shows the off-ridge mooring site for 2000. Depth scale in metres applies to the lower panel.

away from the ridge crest ($L \approx 10$ km is the ridge width), dispersion is still sufficient to transport larvae well away from the ridge into the inhospitable deep ocean. This enhanced cross-axis dispersion by bottom-intensified oscillatory currents above ridge crests presumably carries vent larvae beyond the influence of transport by along-ridge mean currents, regardless of how they are formed. Therefore, despite its obvious appeal, the concept that mean along-axis currents above the ridge crest are the primary factor in successful larval colonization is unlikely.

Contrary to the situation above the ridge crest, oscillatory currents within axial valleys (in the case of slow- and intermediate-spreading regions) or summit grabens (in the case of fast-spreading regions) can be strongly attenuated and rectified by the steep topography. Flow attenuation and rectification increase the likelihood for successful downstream recruitment by creating a volume of water that is sheltered from off-ridge transport. This effect will be augmented by plume-induced inflow for hydrothermally active rift valleys. Plume-induced inflow (Fig. 4) would facilitate the aggregation of vent larvae in the water column near their source region before their export downstream. Larvae with negative buoyancy relative to water above the ridge crest would presumably sink back into the valley should they become swept upward in a plume. These larvae will tend to remain within the valley, just as large ($5\text{ }\mu\text{m}$ diameter) mineral particles tend to be vertically recycled in hydrothermal plumes if they are not advected more than a few kilometres from their source⁹. On the other hand, larvae with positive buoyancy will rise or remain above the ridge crest where they are susceptible to cross-ridge dispersion. For vent organisms such as the tubeworm *Riftia pachyptila*—whose mean lifespan away from the source region is only about 38 days³—

successful transport to an adjacent ridge segment by the mechanism proposed here is most likely to occur during along-ridge flow events that intermittently over-ride the circulation within the valley. Regional large-scale flow events would ‘flush’ the larvae along the ridge axis, the direction most favourable for recruitment. For larvae transported to the vicinity of newly formed vent fields, the presence of recently established plume-induced inflow may facilitate recruitment to the new habitat.

Because steep walls attenuate oscillatory currents over a broad frequency range, rift valleys and escarpments provide stepping stones or conduits for the along-ridge transport of vent-derived products. Intermittent regional flow events then provide the vehicle for rapid and directed transport of vent larvae to distant vent fields. At present, there are too few velocity data for the global mid-ocean ridge system for us to confirm the general applicability of our circulation-based model. However, we can speculate that segment-to-segment colonization by larvae is most successful for

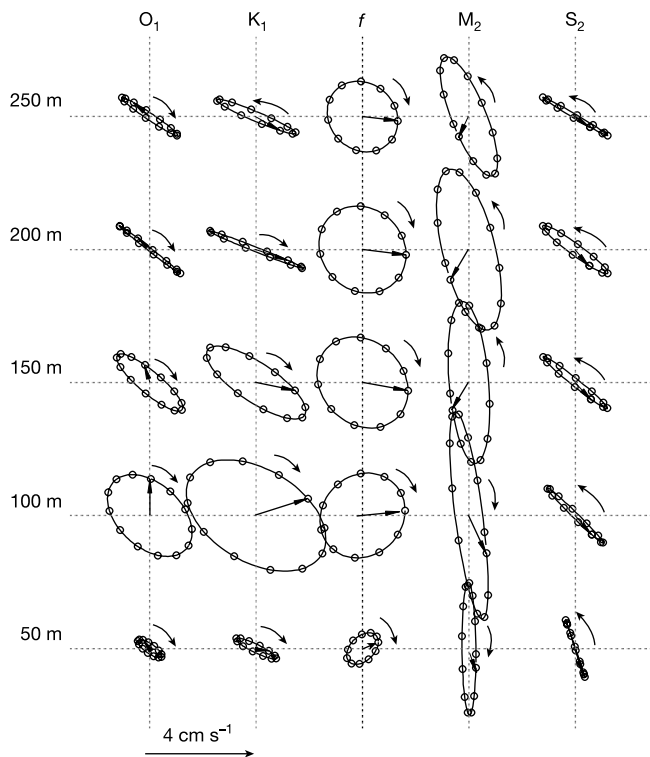


Figure 2 Current ellipses and vectors for the four major tidal constituents and inertial currents (f) for all five depths for the south mooring site in the year 2000. Numbers give instrument depth above bottom. Vectors denote currents at an initial time, t_0 . Small circles mark time steps for the current vectors; arrows outside the ellipse denote the direction of vector rotation. Along-axis is upward in all figures.

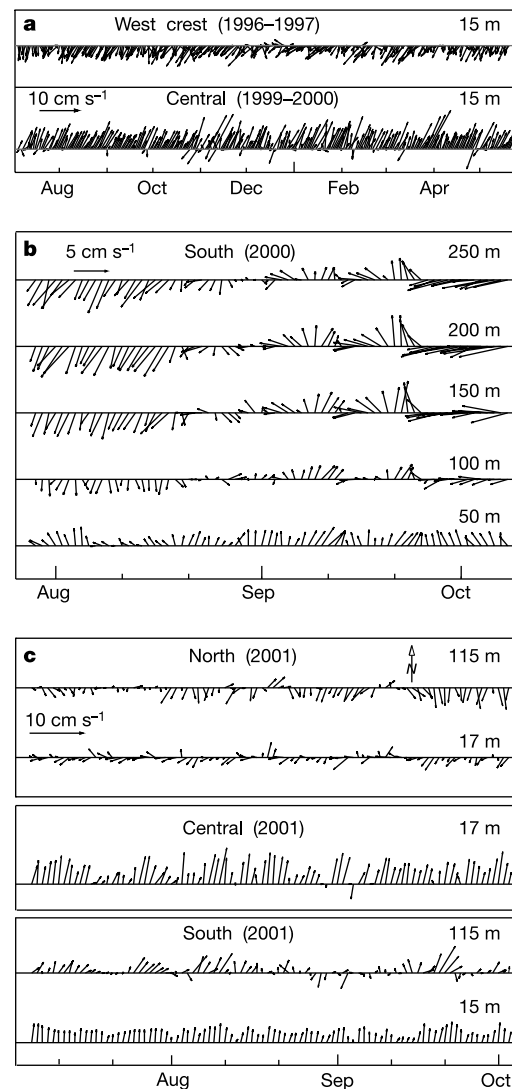


Figure 3 Daily mean current vectors. **a**, West ridge site, June 1996 to July 1997, and the central valley site from June 1999 to May 2000; **b**, the five current meter depths at the south mooring site in year 2000; and **c**, current meters at the north, central and south sites in year 2001. Labels give instrument depth above bottom. North is upward in all figures. Vectors point in the direction of flow. The bottom panel in **a** shows numerous periods of flow reversal within the axial valley.

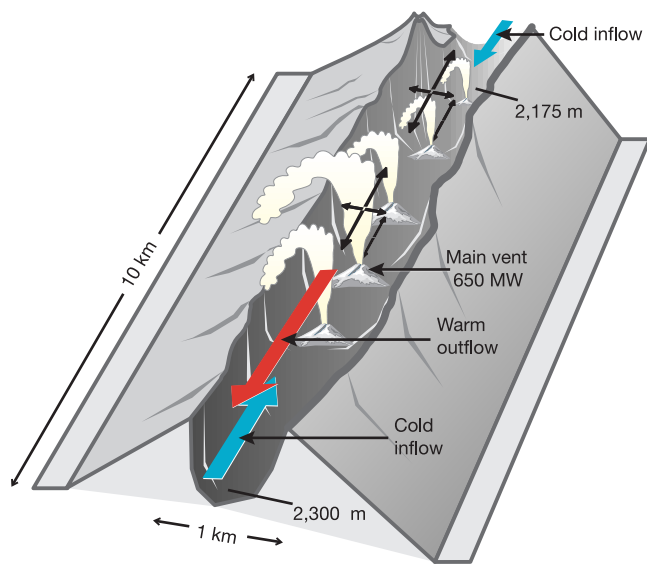


Figure 4 Diagram of the dominant flow within the axial valley of Endeavour ridge. High-frequency oscillatory currents are amplified over the ridge crest (back-and-forth current vectors) but suppressed and rectified within the confines of the valley. Mean background currents carry plume water to the southwest above the ridge crest whereas plume-induced currents transport cold ambient water into the valley below about 75 m elevation. Strongest inflow occurs at the southern end of the valley. Venting dissipates about 1,500 MW, with 650 MW from the Main Endeavour Field alone. Plume sizes are roughly proportional to thermal outputs from the five vent fields.

intermediate-spreading ridges (such as Juan de Fuca ridge), where segments are relatively close together and contain active hydrothermal vent fields in well-defined axial valleys. We further speculate that there is significant colonization success for fast-spreading ridges (such as the East Pacific Rise), where rift valleys are shallower (about 10 m) but more continuous (about 100 km). Possibly less conducive to colonization are the rift valleys of slow-spreading ridges (such as the Mid-Atlantic Ridge) where numerous fracture zones, transform off sets, and large silled valleys can break up ridge segments¹⁰. Major breaks in ridge segments may isolate vent organisms unless the large-scale ocean circulation is predominantly along-ridge. On a geological timescale, our findings imply that periods of enhanced ridge building and associated hydrothermal venting are periods of reduced biogeographic diversity whereby vent larvae are distributed more readily along crests of the global ridge system, and vice versa. Regardless of which processes predominate, our findings strongly express the need for detailed current measurements in any study investigating the dispersion and colonization of vent fauna within the global ridge system. □

Methods

Moored current-meter observations

Current meters were moored 15–17 m above bottom on the west ridge crest from June 1996 to July 1997, near the Main (central) Endeavour Field from June 1999 to May 2000, and at three sites along the axial valley in 2001 (Fig. 1). A single string of five current meters at 50-m depth increments was deployed near the main field from July to October 2000. The latter provides high vertical resolution of the flow in the central valley whereas the former provides reasonably high along-ridge resolution of the flow within the valley.

An empirical model of plume-driven circulation

As proposed for segments of the Mid-Atlantic Ridge¹¹, it is feasible that enhanced diapycnal mixing over the rough topography of the axial valley is a primary cause for the intensified, unidirectional background flow we observed within the south-central portion of the axial valley. However, given the widespread extent of vigorous hydrothermal venting within the valley and observational evidence that the inflow is confined below the base of the neutrally buoyant plumes, it is more likely that the inflow is a dynamic response to the

turbulent entrainment of cold (2 °C) ambient bottom water by plumes rising from superheated (350 °C) and low-temperature (10–25 °C) vent sources. This explanation becomes even more compelling if one examines conservation of fluid mass. Specifically, buoyant plumes rising from the major hydrothermal vent fields located within the southern sector of the valley entrain approximately 5,000 times the mass of the source waters exiting the vent orifices by the time they reach their level of neutral buoyancy¹². Roughly 40% of this entrainment occurs within the first 75 m of rise, well within the confines of the valley. The vertical flux of fluid mass, Q_V , at 75 m elevation is then related to the vertical flux of mass, Q_S , from the bottom sources through the relation

$$Q_V = (0.4)5,000Q_S \quad (1)$$

To support this entrainment, fluid must enter at the ends of the axial valley, particularly from the deeper, more hydrothermally active southern end of the valley. Our observations show that the mean inflow at the southern end of the valley, v_S , is about four times the mean inflow at the northern end, v_N . Thus, by mass conservation

$$Q_V \approx 1.25Av_S \quad (2)$$

where A ($\approx 40 \times 10^3 \text{ m}^2$) is the cross-sectional area of the valley up to an elevation of 75 m. The recent estimate of about $650 \pm 100 \text{ MW}$ for the heat flux for the Main Endeavour Field¹³ (which is larger than previous estimates of about $340 \text{ MW}^{14,15}$) indicates that the main field contributes approximately 40% of the total heat flux $F_S \approx 1,500 \text{ MW}$ for the ridge segment. (Segment scale estimates for Endeavour ridge^{16–18} range from 1,000 to 3,000 MW.) Relating the mass flux of the segment sources to the corresponding heat flux gives

$$Q_S \approx F_S / \rho C_p \Delta\theta \quad (3)$$

where $\rho C_p \approx 4.3 \times 10^6 \text{ J K}^{-1} \text{ m}^{-3}$ is the heat capacity per unit mass for water and $\Delta\theta \approx 350^\circ\text{C}$ is the typical temperature contrast between vent fluids and the ambient water. Solving equations (1) to (3) for the mean inflow, v_S , at the southern end of the valley yields

$$v_S \approx \frac{(0.4)5,000F_S}{1.25A\rho C_p \Delta\theta} \approx \frac{(0.4)5,000 \times 1500 \times 10^6 \text{ J s}^{-1}}{(1.25)(4 \times 10^4 \text{ m}^2)(4.3 \times 10^6 \text{ J K}^{-1} \text{ m}^{-3})350 \text{ K}} \approx 4 \text{ cm s}^{-1}$$

The plume-induced inflow at the northern end of the valley is then $v_N \approx 1 \text{ cm s}^{-1}$. Our estimates for v_S and v_N become almost identical with observed background currents at their respective ends of the valley if we account for the fact that inflows are typically offset by a southward regional current of $1\text{--}2 \text{ cm s}^{-1}$.

One study observed a deep ($>3,010 \text{ m}$ depth) influx $Q_U = 5.30 \pm 0.55 \times 10^3 \text{ m}^3 \text{ s}^{-1}$ within a 50-km-long, silled segment of the Mid-Atlantic Ridge at 29°N but provided no mechanistic link between the flux and the 275-MW output from hydrothermal vent fields contained within the valley¹⁰. However, if we assume that the valley has $N = 10$ vent sites, each having a surface area $A = 0.03 \text{ m}^2$, plume exit velocities of $w = 1 \text{ m s}^{-1}$, and a dilution factor of 7,500:1 (ref. 19), we obtain a vertical flux of entrained fluid $Q_V = NAw7,500 = 2.25 \times 10^3 \text{ m}^3 \text{ s}^{-1}$, about 40% of the observed value, Q_U . Thus, thermal buoyancy effects may also be contributing to the deep inflow at this segment of the Mid-Atlantic Ridge.

Received 7 February; accepted 2 June 2003; doi:10.1038/nature01824.

1. Tunnicliffe, V. The biology of hydrothermal vents: Ecology and evolution. *Oceanogr. Mar. Biol. Ann. Rev.* **29**, 319–407 (1991).
2. van Dover, C. L., German, C. R., Speer, K. G., Parson, L. M. & Vrijenhoek, R. C. Evolution and biogeography of deep-sea vents and seep invertebrates. *Science* **295**, 1253–1257 (2002).
3. Marsh, A. G., Mullineaux, L. S., Young, C. M. & Manahan, D. T. Larval dispersal potential of the tubeworm *Riftia pachyptila* at deep-sea hydrothermal vents. *Nature* **411**, 77–80 (2001).
4. Kelley, D. S., Delaney, J. R. & Lilley, M. D. Vent fluid distribution and evolution along the Endeavour segment, Juan de Fuca Ridge. *EOS Trans. AGU* **82**, 47, F612 (2001).
5. Stahr, F. R., McDuff, R. E., Yoerger, D. R., Bradley, A. M. & Nakamura, K. Heat flux measurements at the Main Endeavour vent field, Juan de Fuca Ridge. *Eos Trans. AGU* (Fall Mtg Suppl. Abstr.) **81**(48), (2000) O5521-03.
6. Cannon, G. A. & Thomson, R. E. Characteristics of a 4-day oscillation trapped by the Juan de Fuca Ridge. *Geophys. Res. Lett.* **23**, 1613–1616 (1996).
7. Mihaly, S. E., Thomson, R. E. & Rabinovich, A. B. Evidence for nonlinear interaction between internal waves of inertial and semi-diurnal frequency. *Geophys. Res. Lett.* **25**, 1205–1208 (1998).
8. Lavelle, J. W. & Cannon, G. A. On subinertial oscillations trapped by the Juan de Fuca Ridge, northeast Pacific. *J. Geophys. Res.* **106**, 31099–31116 (2001).
9. German, C. R. & Sparks, R. S. J. Particle recycling in the TAG hydrothermal plume. *Earth Planet. Sci. Lett.* **116**, 129–134 (1993).
10. Murton, B. J., Redbourn, L. J., German, C. R. & Baker, E. T. Sources and fluxes of hydrothermal heat, chemicals and biology within a segment of the Mid-Atlantic Ridge. *Earth Planet. Lett.* **171**, 301–317 (1999).
11. Thurnherr, A. M., Richards, K. J., German, C. R., Lane-Serff, G. F. & Speer, K. G. Flow and mixing in the Rift Valley of the Mid-Atlantic Ridge. *J. Phys. Oceanogr.* **32**, 1763–1778 (2002).
12. McDuff, R. E. In *Physical, Chemical, Biological, and Geological Interactions within Seafloor Hydrothermal Systems* (eds Humphris, S. E., Zierenberg, R. A., Mullineaux, L. S. & Thomson, R. E.) 357–368 (American Geophysical Union Monograph, Washington, 1995).
13. Veirs, S. R. *Heat Flux and Hydrography at a Submarine Volcano: Observations and Models of the Main Endeavour Vent Field in the Northeast Pacific* PhD thesis, Univ. Washington (2003).
14. Bemis, K. G., von Herzen, R. P. & Mottl, M. J. Geothermal heat flux from hydrothermal plume on the Juan de Fuca Ridge. *J. Geophys. Res.* **98**, 6351–6369 (1993).
15. Ginster, U., Mottl, M. J. & von Herzen, R. P. Heat flux from black smokers on the Endeavour and Clift segments, Juan de Fuca Ridge. *J. Geophys. Res.* **99**, 4937–4950 (1994).
16. Baker, E. T. & Massoth, G. J. Characteristics of hydrothermal plumes from two vent fields on the Juan de Fuca Ridge, northeast Pacific Ocean. *Earth Planet. Sci. Lett.* **85**, 59–73 (1987).

17. Rosenberg, N. D. *et al.* Estimation of heat and chemical fluxes from a seafloor hydrothermal vent field using radon measurements. *Nature* **334**, 604–607 (1988).
18. Thomson, R. E., Delaney, J. R., McDuff, R. E., Janecky, D. R. & McClain, J. S. Physical characteristics of the Endeavour Ridge hydrothermal plume during July 1988. *Earth Planet. Sci. Lett.* **111**, 141–154 (1992).
19. Lupton, J. E., Delaney, J. R., Johnson, H. P. & Tivey, M. K. Entrainment and vertical transport of deep-ocean water by buoyant hydrothermal plumes. *Nature* **316**, 621–623 (1985).

Acknowledgements We thank the IOS Mooring Group of T. Juhász and D. Spear for successful deployment and recovery of all instruments; K. Conley and D. Andrie of the Oceans Division of Fisheries and Oceans Canada for helping fund the moorings; A. Lee for assisting with the data processing; and P. Kimber for assisting with the figures. This work was partially supported by a Natural Sciences and Engineering Research Council (NSERC) grant and National Science Foundation (NSF) grant.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.E.T. (Thomsonr@pac.dfo-mpo.gc.ca).

Evolutionary capacitance as a general feature of complex gene networks

Aviv Bergman^{1,2} & Mark L. Siegal¹

¹Department of Biological Sciences, Stanford University, Stanford, California 94305-5020, USA

²Center for Computational Genetics and Biological Modeling, Stanford University, Stanford, California 94305-5020, USA

An evolutionary capacitor buffers genotypic variation under normal conditions, thereby promoting the accumulation of hidden polymorphism. But it occasionally fails, thereby revealing this variation phenotypically¹. The principal example of an evolutionary capacitor is Hsp90, a molecular chaperone that targets an important set of signal transduction proteins. Experiments in *Drosophila* and *Arabidopsis* have demonstrated three key properties of Hsp90: (1) it suppresses phenotypic variation under normal conditions and releases this variation when functionally compromised; (2) its function is overwhelmed by environmental stress; and (3) it exerts pleiotropic effects on key developmental processes^{1,2}. But whether these properties necessarily make Hsp90 a significant and unique facilitator of adaptation^{1–10} is unclear. Here we use numerical simulations of complex gene networks, as well as genome-scale expression data from yeast single-gene deletion strains, to present a mechanism that extends the scope of evolutionary capacitance beyond the action of Hsp90 alone. We illustrate that most, and perhaps all, genes reveal phenotypic variation when functionally compromised, and that the availability of loss-of-function mutations accelerates adaptation to a new optimum phenotype. However, this effect does not require the mutations to be conditional on the environment. Thus, there might exist a large class of evolutionary capacitors whose effects on phenotypic variation complement the systemic, environment-induced effects of Hsp90.

Several studies on other genes than that encoding Hsp90 have also reported increased phenotypic variance in mutant strains^{11,12}. These results have been taken as evidence that wild-type organisms are buffered, or ‘canalized’, against environmental and genetic variation¹³. We previously demonstrated that canalization does not require a dedicated mechanism (such as that provided by chaperone proteins), but instead arises as an emergent property of complex developmental-genetic networks that reach a steady state

of gene expression¹⁴. We therefore proposed that a loss of buffering might be induced by compromising the function of an evolved network, namely by removing the function of an arbitrary gene in the network. To test this hypothesis, we use simulations of gene networks and ask how an arbitrary null (‘knockout’) mutation affects the expression of other genes. Does a network with a single gene knocked out exhibit a greater range of phenotypes, across environmental conditions or genetic backgrounds, than does the wild-type network from which it was derived? If so, the knocked-out gene has the potential to be an evolutionary capacitor.

First we consider the effect of knockouts in different genetic backgrounds. We use a representation of a gene network in which each of N genes produces a factor that is capable of influencing the expression of each other gene as well as itself^{14,15} (see Fig. 1 and Supplementary Information; $N = 10$ for all simulations presented here). An $N \times N$ matrix encapsulating the regulatory relationships is considered the ‘genotype’ of an individual. Unlike standard quantitative-genetic and population-genetic frameworks, this representation lends itself to the study of developmental buffering, because development is explicitly modelled as the progression from an initial gene-expression state to a state of gene-expression equilibrium. The equilibrium gene-expression profile is considered to be the ‘phenotype’, and those genotypes that do not produce an equilibrium state are considered lethal (see refs 14 and 15 for a justification of this definition of lethality). A knockout of a gene is represented by zeroing the corresponding row and column of the regulatory matrix.

To study a population of individuals that differ in genotype but, for the most part, have the same phenotype, we perform 100 evolutionary simulations in which a randomly generated, non-lethal individual founds a population of size 500 that is allowed to evolve, with mutation of the elements of the regulatory matrix, for 400 generations. During evolution there is strong selection both against lethal disruptions in gene-expression equilibrium and for the phenotype of the founder individual, as described previously¹⁴. Setting the founder’s phenotype as optimal in this way ensures that the range of phenotypes in the final population of 500 individuals at generation 400 is very narrow, thus permitting comparison with the range of phenotypes seen in the knockout derivatives of each of the 500 individuals. Out of 1,000 such comparisons (100 simulations $\times$ 10 gene knockouts), only 4 showed higher phenotypic variation for the wild-type population than for its knockout derivatives ($P \approx 0$, sign test; for a detailed analysis see Supplementary Information). Results of a typical simulation are shown in

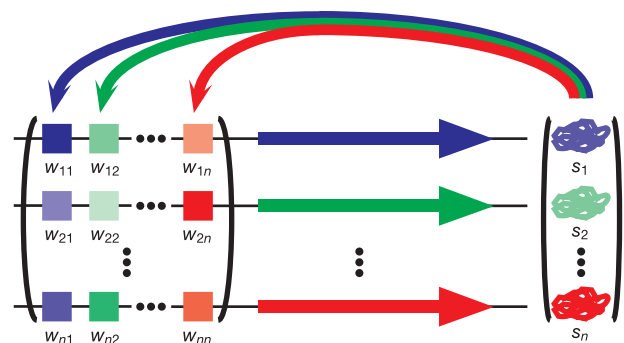


Figure 1 Representation of a gene network. Each gene (horizontal arrow) is regulated by the products of the other genes by means of upstream enhancer elements (boxes). The strength and direction of regulation (depicted as different colour saturation levels) are a function of both the regulatory element and the abundance of its corresponding gene product. Genotype is represented as the matrix, W , of regulatory interactions, and phenotype is the vector, $\hat{S}$, of gene-product levels at equilibrium.

Fig. 2. The range of phenotypes in the final population is indeed quite narrow, with nearly all individuals having the optimum phenotype. In striking contrast, the range of phenotypes seen when each individual has a given gene knocked out is quite broad. As shown in Fig. 2, the exact proportion of genetic backgrounds on which a knockout is lethal, and also the distribution of non-lethal phenotypes, depend on which gene is knocked out, but the evolved population clearly harbours genetic variation that is not expressed phenotypically unless a gene is knocked out. Indeed, each knockout reveals phenotypic variation that is not seen in the wild-type population. Interestingly, the most frequent phenotypic class is generally that corresponding to the optimum phenotype, indicating that the evolved networks have some robustness even to these major-effect mutations. Also noteworthy is that the property of buffering genetic variation seems to be relatively independent of the exact network topology, because we observed it in all simulations performed as described here and also in similar simulations in which the network of the population founder was generated not from a Poisson distribution of connections but rather from a power-law distribution of outgoing connections, creating a so-called 'scale-free' topology, which has been inferred for the gene-regulation network of yeast¹⁶ and other biological networks¹⁷. We therefore conclude that arbitrary genes in complex gene networks have the property of buffering genetic variation and therefore have the potential to act as evolutionary capacitors.

As Hsp90's effects are dependent on the environment, and as the relationship between genetic and environmental variation features prominently in theoretical considerations of canalization^{10,18}, we also investigated the effect of knockouts on individuals subject to environmental 'noise'. Noise is introduced into the model used above by the random perturbation of an individual's initial gene-expression state. The variance in phenotypes induced by variation in initial conditions is then measured for an individual and for each single-gene knockout derivative of that individual. In addition to this comparison between wild type and knockout, we also compare individuals 'before' and 'after' evolution, to assess the effect of genetic canalization on the response to environmental perturbation.

The evolutionary simulations are performed as above (but without selection for an optimum phenotype), and from each the population founder is assayed as a before-evolution individual, and the most canalized (that is, the least sensitive to mutations) descendant present in the final generation is assayed as an after-evolution individual. Thus, each simulation contributes to four categories of individuals tested for their sensitivity to initial conditions: wild type before evolution, knockout before evolution, wild type after evolution, and knockout after evolution.

Both for before-evolution and for after-evolution individuals, the average variance in gene expression across all genes tends to be lower for a wild-type individual subjected to environmental noise than for its knockout derivatives under the same conditions. The amplitude of noise required to induce non-zero variance differs from individual to individual, so for each comparison between wild type and knockout we assayed a range of noise levels and used the lowest level at which the wild-type and knockout individuals had different variance. By this measure, of 500 wild-type before-evolution individuals, 434 displayed lower gene-expression variance than in the corresponding knockouts, and 66 displayed higher variance than in the corresponding knockouts. The results are similar for after-evolution individuals, with 479 lower in the wild type and 21 higher in the wild type. Notice, however, that for the after-evolution individuals there is a greater proportion of wild-type individuals with lower variance than the corresponding knockouts (G -test with Williams correction, $P = 2.6 \times 10^{-7}$). This could be a result of either a lower sensitivity to initial conditions for the after-evolution wild-type individuals or a greater sensitivity to initial conditions for the knockout derivatives of the after-evolution individuals. The former seems to be true. Of 500 pairs of before-evolution and after-evolution individuals, 390 displayed higher gene-expression variance in the before-evolution individual than in the after-evolution individual; for the corresponding knockouts, 364 displayed higher gene-expression variance before evolution than after evolution. Thus, not only are evolved individuals less sensitive to initial conditions, but their knockout derivatives are also less sensitive (no significant difference between wild type and

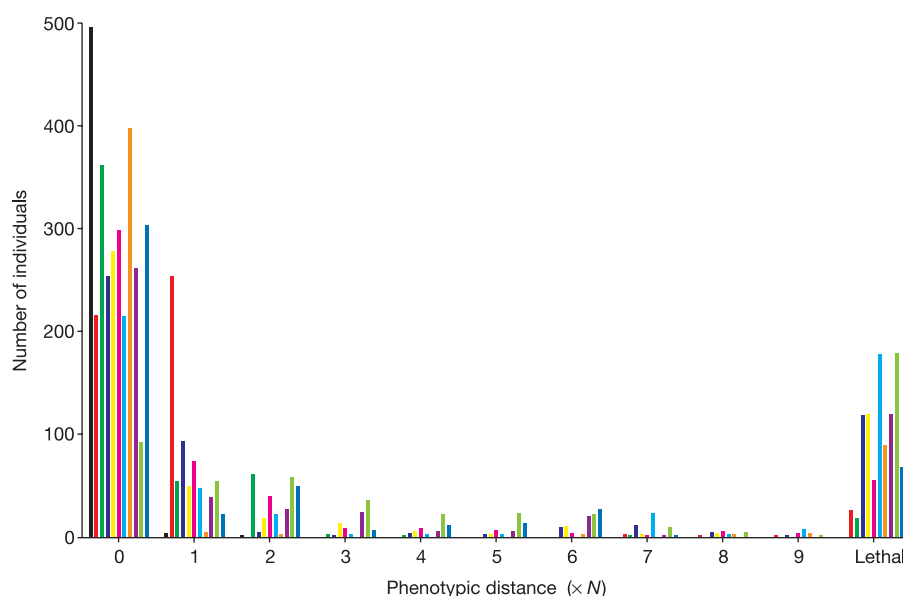


Figure 2 Greater phenotypic variation in single-gene knockouts than in the wild-type networks from which they derive. The histogram shows results of a typical simulation, in which a population of 500 individuals evolved for 400 generations under strong stabilizing selection. Black bars show the distribution of phenotypes in the final population, as the phenotypic distance (defined previously¹⁴) to the selected optimum (phenotypic

distances are multiplied by N , the number of genes in the network (10 for wild type, 9 for knockouts)). Each set of bars of the same colour shows the distribution of phenotypes resulting from introducing a given single-gene knockout into each of the 500 members of the final population. The rightmost group of bars shows the number of lethal genotypes for each knockout.

knockouts in this respect, $P = 0.056$). We conclude that the evolved individuals, although not selected directly to become less sensitive to variation in initial conditions, have achieved this property. It is also clear that knockout mutations significantly increase the sensitivity to initial conditions of both before-evolution and after-evolution individuals. These results were not noticeably changed when the evolved individuals were also subject to selection for a phenotypic optimum (not shown), similar to what was seen previously for the evolution of canalization itself⁴.

At this point one might ask whether there is any corroborating experimental evidence that knockout mutations tend to increase variation in the expression of other genes. Indeed there is: a systematic project has been undertaken to delete in turn each gene in the yeast *Saccharomyces cerevisiae*; each knockout strain is then assayed for the expression levels of all remaining genes as well as for its growth rate relative to the other strains^{19–21}. We have analysed the available data from this project for evidence that the buffering of other genes' expression levels is disrupted when one gene is knocked out. The yeast strains are isogenic, so it is buffering against environmental, not genetic, variation that is being probed. However, this distinction is not likely to be important because it has been argued that environmental and genetic buffering are likely to be achieved by the same mechanisms^{10,22}, and because our simulations above indicate that the evolution of genetic canalization makes individuals less sensitive to environmental variation as well. In an initial analysis of the yeast data²⁰ a measure called a 'scale factor' was introduced, which was calculated for each gene using data from 63 wild-type control experiments and captures the degree of biological variability in expression for that gene. When we do the same calculation using instead the expression data from 53 knockouts, the resulting value for each gene tends to be higher than its scale factor; this is true even for those genes that do not show any evidence of being a regulatory target of any one of the knockouts ($P < 0.0001$, Friedman's test). Thus, deleting a gene tends to increase the variation in expression of other genes. Moreover, the increase in gene-expression variation is linked to phenotypic variation, in that the magnitude of revealed variation is negatively correlated with fitness in the standard growth condition: knockout strains with lower fitness tend to be those with greater variation in the expression levels of the remaining non-target genes and vice versa (product-moment correlation $r = -0.1973$, significant at $P = 0.0278$ by non-parametric permutation test).

So far we have shown that removing the function of an arbitrary gene tends to increase the variance in expression of other genes in its network, ultimately increasing the phenotypic variance. The key question now is whether this effect has important evolutionary consequences. Because one of the special features of Hsp90 is the environmental sensitivity of its function, a related question is whether only genes with conditional alleles can serve as evolutionary capacitors. To answer these questions we perform evolutionary simulations as before, but with an additional mutational process by which a gene's activity can be knocked out and, at a lower rate, restored to function. We allow populations to evolve for 400 generations under strong selection for an optimum phenotype, then shift to a new optimum at which the equilibrium expression states of three of the 10 network genes flip from on to off or vice versa. We simulated 200 founder populations with their corresponding initial and shifted optima, and allowed each population to evolve under two models: with and without the knockout mutation process. Populations evolving with the occurrence of knockout mutations reached the new optimum an average of 77 generations faster than those with no knockouts (314 generations with knockouts compared with 391 generations without; $P = 0.0030$, Friedman's test), indicating both that evolutionary capacitance does

indeed speed adaptation to a new phenotypic optimum and that mutations that disrupt buffering need not be conditional on the environment that enforces the new optimum. Interestingly, if the population is shifted to its new optimum immediately after its founding (instead of after 400 generations), the average number of generations to reach the new optimum is not significantly dependent on the presence or absence of knockout mutations (379 generations with knockout compared with 461 generations without; $P = 0.066$). This indicates that the accumulation of allelic variation during the evolution of a canalized genetic system might be a prerequisite for the adaptive advantage of loss-of-function mutations. The interplay between accumulated, but hidden, variation and its knockout-induced revelation could provide a mechanistic link to a theoretical model that predicts that selection towards a new optimum promotes decanalization²³.

We showed previously that genetic canalization, or phenotypic insensitivity to mutation, is an emergent property of complex gene networks, in that such networks evolve greater insensitivity to mutation without direct selection for this property and even without selection for an optimum phenotype¹⁴. Here we have extended this result by showing that both environmental canalization and genetic canalization break down when an arbitrary gene is deleted and that the availability of knockout mutations increases the rate of adaptation to a new phenotypic optimum. These results indicate that evolutionary capacitance is not limited to dedicated buffering mechanisms such as Hsp90 but might also be exhibited by many other genes, by virtue of their being part of a complex genetic network. This finding is consistent with the previously mentioned studies^{11,12} demonstrating increased phenotypic variation in mutants compared with the wild type. It is also consistent with theoretical models that predict an increased phenotypic effect of new mutations after mutating a gene with epistatic relationships to other genes^{24,25}, although such models do not predict canalization to exist unless selectively favoured.

Hsp90 has three features that make it appealing as a candidate evolutionary capacitor: its effect of revealing phenotypic variation when its function is impaired or overwhelmed, its pleiotropic effects on several key developmental processes, and the environmental sensitivity of its buffering capacity. We have presented an alternative mechanism with partly overlapping features. The revelation of phenotypic variation is central to the proposed mechanism, as it is to Hsp90, and we have shown that this property is expected to be shared by many genes. These genes will differ in their degree of pleiotropy, but it is likely that collectively they affect a range of phenotypes as broad as, or broader than, that affected by Hsp90; even individually some might approach the level of pleiotropy of Hsp90, because key genetic pathways tend to be reused during development. Sensitivity to environmental conditions is also not unique to Hsp90, because conditional (for example, temperature-sensitive) alleles are fairly commonly recovered in genetic screens. Conditionality might also be achieved by other mechanisms; for example, a gene's function might be eliminated by the insertion of a transposable element and then later restored by excision of the element²⁶. Beyond this, our results suggest that conditionality is not an absolute requirement of an evolutionary capacitor, because in our simulations the existence of non-conditional knockouts accelerated adaptation to a new phenotypic optimum. Why are such knockouts tolerated, let alone favoured during adaptation to a new optimum? The answer might lie in robustness itself. That is, some genes might exist solely to stabilize the function of a network; these genes would seem expendable in the laboratory, yet would be selectively maintained in nature until their loss facilitated adaptation. The existence in model organisms of many gene knockouts without apparent phenotypic effect²⁷ lends plausibility to this idea. It remains to be seen what types of mutations other than knockouts can serve to decanalize a genetic system. In particular, it will be important to explore the effects of partial loss-of-function

mutations, perhaps in the context of a diploid model analogous to the haploid one presented here.

Hsp90 might be a member of a fairly large class of genes with the ability to serve as evolutionary capacitors, most of which might have more subtle effects than Hsp90 has. The existence of this complementary set of capacitors means that adaptation by means of the release of canalized variation might be possible even under conditions that do not involve chronic environmental stress, such as sexual selection and predator–prey ‘arms races’. The proposed capacitance mechanism could be tested by the analysis of genotypic and phenotypic data from populations of a laboratory microorganism, such as *S. cerevisiae* or *Escherichia coli*, at successive stages in adaptation to an artificial selection regime. As for natural adaptations, establishing the role of Hsp90 or another potential capacitor in facilitating the evolution of any particular trait will require intensive investigation of the relevant ecological, population-biological and evolutionary-genetic history. □

Methods

Gene networks and their evolution were modelled as described previously¹⁴. The original model, and extensions of it used here, are described fully in Supplementary Information. The yeast gene-expression data are from ref. 20. The statistical tests presented use the same stringent criterion for declaring a gene a regulatory target of the knocked out gene as that used by those authors. To ensure the robustness of our results we repeated the tests with a range of more lenient criteria; in all cases the test results remained significant. Details of our analysis, including a list of microarray experiments used, are also available in Supplementary Information.

Received 26 November 2002; accepted 24 April 2003; doi:10.1038/nature01765.

- Rutherford, S. L. & Lindquist, S. Hsp90 as a capacitor for morphological evolution. *Nature* **396**, 336–342 (1998).
- Queitsch, C., Sangster, T. A. & Lindquist, S. Hsp90 as a capacitor of phenotypic variation. *Nature* **417**, 618–624 (2002).
- Cossins, A. Cryptic clues revealed. *Nature* **396**, 309–310 (1998).
- Dickinson, W. J. & Seger, J. Cause and effect in evolution. *Nature* **399**, 30 (1999).
- McLaren, A. Too late for the midwife toad: Stress, variability and Hsp90. *Trends Genet.* **15**, 169–171 (1999).
- Wagner, G. P., Chiu, C.-H. & Hansen, T. F. Is Hsp90 a regulator of evolvability? *J. Exp. Zool.* **285**, 116–118 (1999).
- Rutherford, S. L. From genotype to phenotype: buffering mechanisms and the storage of genetic information. *BioEssays* **22**, 1095–1105 (2000).
- Pigliucci, M. Developmental genetics: buffer zone. *Nature* **417**, 598–599 (2002).
- Mitchell-Olds, T. & Knight, C. A. Chaperones as buffering agents? *Science* **296**, 2348–2349 (2002).
- Meiklejohn, C. D. & Hartl, D. L. A single mode of canalization. *Trends Ecol. Evol.* **17**, 468–473 (2002).
- Scharloo, W. Canalization: Genetic and developmental aspects. *Annu. Rev. Ecol. Syst.* **22**, 65–93 (1991).
- Gibson, G. & Wagner, G. Canalization in evolutionary genetics: A stabilizing theory? *BioEssays* **22**, 372–380 (2000).
- Waddington, C. H. Canalization of development and the inheritance of acquired characters. *Nature* **150**, 563–565 (1942).
- Siegal, M. L. & Bergman, A. Waddington's canalization revisited: Developmental stability and evolution. *Proc. Natl Acad. Sci. USA* **99**, 10528–10532 (2002).
- Wagner, A. Does evolutionary plasticity evolve? *Evolution* **50**, 1008–1023 (1996).
- Featherstone, D. E. & Broadie, K. Wrestling with pleiotropy: Genomic and topological analysis of the yeast gene expression network. *BioEssays* **24**, 267–274 (2002).
- Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N. & Barabási, A.-L. Hierarchical organization of modularity in metabolic networks. *Science* **297**, 1551–1555 (2002).
- Wagner, G. P., Booth, G. & Bagheri-Chaichian, H. A population genetic theory of canalization. *Evolution* **51**, 329–347 (1997).
- Winzler, E. A. *et al.* Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285**, 901–906 (1999).
- Hughes, T. R. *et al.* Functional discovery via a compendium of expression profiles. *Cell* **102**, 109–126 (2000).
- Giaever, G. *et al.* Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418**, 387–391 (2002).
- Stearns, S. C., Kaiser, M. & Kawecki, T. J. The differential genetic and environmental canalization of fitness components in *Drosophila melanogaster*. *J. Evol. Biol.* **8**, 539–557 (1995).
- Rice, S. H. The evolution of canalization and the breaking of von Baer's laws: Modeling the evolution of development with canalization. *Evolution* **52**, 647–656 (1998).
- Wagner, G. P. & Mezey, J. Modeling the evolution of genetic architecture: A continuum of alleles model with pairwise A × A epistasis. *J. Theor. Biol.* **203**, 163–175 (2000).
- Hansen, T. F. & Wagner, G. P. Modeling genetic architecture: A multilinear theory of gene interaction. *Theor. Popul. Biol.* **59**, 61–86 (2001).
- Edwards, R. J. & Brookfield, J. F. Y. Transiently beneficial insertions could maintain mobile DNA sequences in variable environments. *Mol. Biol. Evol.* **20**, 30–37 (2003).
- Tautz, D. A genetic uncertainty problem. *Trends Genet.* **16**, 475–477 (2000).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank B. Baker, D. Hartl, J. Hermisson, D. Kennedy, J. Masel, C. Meiklejohn, D. Petrov, G. Wagner, C. Yanofsky and E. Zuckerkandl for helpful discussions. This study was supported by the Center for Computational Genetics and Biological Modeling, Stanford University. M.L.S. was supported by a National Institutes of Health National Research Service Award Individual Postdoctoral Fellowship, and thanks B. Baker for his support. A.B. thanks the Paul G. Allen Charitable Foundation for its continual support.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.L.S. (mlsiegel@stanford.edu).

Organization of cell assemblies in the hippocampus

Kenneth D. Harris, Jozsef Csicsvari*, Hajime Hirase, George Dragoi & György Buzsáki

Center for Molecular and Behavioral Neuroscience, Rutgers, The State University of New Jersey, 197 University Avenue, Newark, New Jersey 07102, USA

* Present address: MRC Anatomical Neuropharmacology Unit, Mansfield Road, Oxford OX1 3TH, UK

Neurons can produce action potentials with high temporal precision¹. A fundamental issue is whether, and how, this capability is used in information processing. According to the ‘cell assembly’ hypothesis, transient synchrony of anatomically distributed groups of neurons underlies processing of both external sensory input and internal cognitive mechanisms^{2–4}. Accordingly, neuron populations should be arranged into groups whose synchrony exceeds that predicted by common modulation by sensory input. Here we find that the spike times of hippocampal pyramidal cells can be predicted more accurately by using the spike times of simultaneously recorded neurons in addition to the animals location in space. This improvement remained when the spatial prediction was refined with a spatially dependent theta phase modulation^{5–8}. The time window in which spike times are best predicted from simultaneous peer activity is 10–30 ms, suggesting that cell assemblies are synchronized at this timescale. Because this temporal window matches the membrane time constant of pyramidal neurons⁹, the period of the hippocampal gamma oscillation¹⁰ and the time window for synaptic plasticity¹¹, we propose that cooperative activity at this timescale is optimal for information transmission and storage in cortical circuits.

In sensory brain regions, the temporal pattern of spikes can correlate precisely with the time course of an external stimulus^{12–14}. In high-level structures, however, neural responses are often more variable than is expected from sensory control¹⁵. Is this variability simply noise¹⁶, or does it reflect the operation of internal, non-sensory processes? In the hippocampus, the timing of pyramidal cell spikes with respect to a clock (‘theta’) rhythm is correlated with the animal's location in space^{5,6}. This timing does not reflect the occurrence of external sensory events precisely timed with respect to the theta rhythm, but rather must arise because of dynamics internal to the brain, a conclusion that is reinforced by the existence of similar phenomena during non-spatial behaviours⁷.

We have investigated the hypothesis that hippocampal neurons are organized in time into ‘cell assemblies’ whose activity can reflect both external sensory input and internal cognitive processes^{2,17}. A signature of assembly organization is the existence of anatomically distributed groups of neurons whose activity is synchronized more than is predicted by common sensory modulation. A second postulated signature is that, although individual neurons may participate in many assemblies, not every possible combination of

cells comprises a cell assembly. The latter feature should be reflected by a statistical preference in the probability with which a neuron joins its various peers in synchronous firing.

The activity of simultaneously monitored pyramidal cells, with neurons arranged according to their physical position within the CA1 region, showed no apparent spatio-temporal patterns^{18,19} (Fig. 1a, b). When the units were rearranged so that synchronously firing cells were displayed near each other, however, the patterns were often suggestive of a cell assembly organization, with different sets of cells repeatedly showing synchronous activity at different times (Fig. 1c).

To verify that the visualized assembly organization was not simply due to chance coincidences of spikes, we used a 'peer prediction' method (Fig. 2). Because hippocampal pyramidal cells have place

correlates²⁰, their instantaneous firing probabilities can be predicted from the spatial position of the rat. Under the assumption of a 'firing rate code' for position^{16,21,22}, this is the best prediction that can be made for a cell's activity. By contrast, the cell assembly hypothesis predicts that cells should show synchronous activation beyond that predicted by their common modulation by sensory input. Thus, it should be possible to predict a cell's exact spike times better from the activity of simultaneously recorded assembly members than from spatial position alone.

To predict the spike train of a target cell, we estimated the discharge probability of the target cell at every instant of time (the 'instantaneous intensity function'²³). To predict intensity from position alone, the mean firing rate for the animal's location was determined from the place field (Fig. 2b). To predict from peer activity (Fig. 2a), each peer cell was assigned a 'weight', which was either positive or negative (Methods and Supplementary Information). The activity of positively or negatively weighted cells increased or decreased the predicted intensity of the target cell within a specific time window (the 'peer prediction timescale').

To determine the weights and to quantify the success of prediction, we used a cross-validation method: weights were chosen to maximize the success of prediction on one part of the data (the training set), and the success of prediction was quantified on the remaining part (the test set). The cross-validation method is essential to ensure that the predictions are not due to statistical fluctuations (Supplementary Information). The prediction method is shown in Fig. 2c, d. Typically, positively and negatively weighted cells showed peaks and dips in their cross-correlograms (CCGs) with the target cell (Fig. 2e).

Of the 200 pyramidal cells that met our unit isolation criteria (Methods and Supplementary Information), the spike trains of 189 (95%) were predictable from space, which were therefore classified as 'place cells'. For most place cells ($n = 185$, 98%), adding population activity to the place information further improved spike timing predictability (Fig. 3a). Additional peer predictability increased with the number of simultaneously monitored neurons ($r = 0.55$, $P < 10^{-14}$; Supplementary Information).

By varying the peer prediction timescale, the temporal window within which spike times were best predictable from the population could be determined. For short temporal windows (~ 1 ms) prediction was poor, indicating that spike timing was not coordinated to this accuracy in the population. Prediction improved with longer peer prediction timescales, reaching a maximum predictability of 4.8 bits s^{-1} at 25 ms for a representative neuron (Fig. 3b), as compared with 3.2 bits s^{-1} for prediction from space alone. The distribution of the optimal time windows across the population showed a peak between 10 and 30 ms (Fig. 3c; 95% confidence interval for median, 21–26 ms). These observations suggest that pyramidal neurons are organized into assemblies whose activity synchronizes transiently with a temporal resolution of ~ 25 ms. Positively weighted cells are those with which the target cell frequently joins in an assembly, and negatively weighted cells are those with which the target cell rarely joins in an assembly.

A mechanism that can affect neuronal firing times is the hippocampal theta oscillation²⁴. The timing of spikes within the theta cycle, in addition to the cell's firing rate, is correlated with the rat's position in space^{5–8,25}. It may be argued, therefore, that correlated assembly firing arises from the simultaneous phase modulation of cell populations. We considered two schemes. In the first, neuronal synchronization is due to the dependence of the mean theta phase of each cell on position, but fluctuations in timing about this mean are random and independent between cells. In the second, both mean phase and correlated phase fluctuations contribute to synchronization, in which case the phase–space correlation is only one manifestation of a more general mechanism determining exact spike times.

To compare these possibilities, we refined the prediction of spike

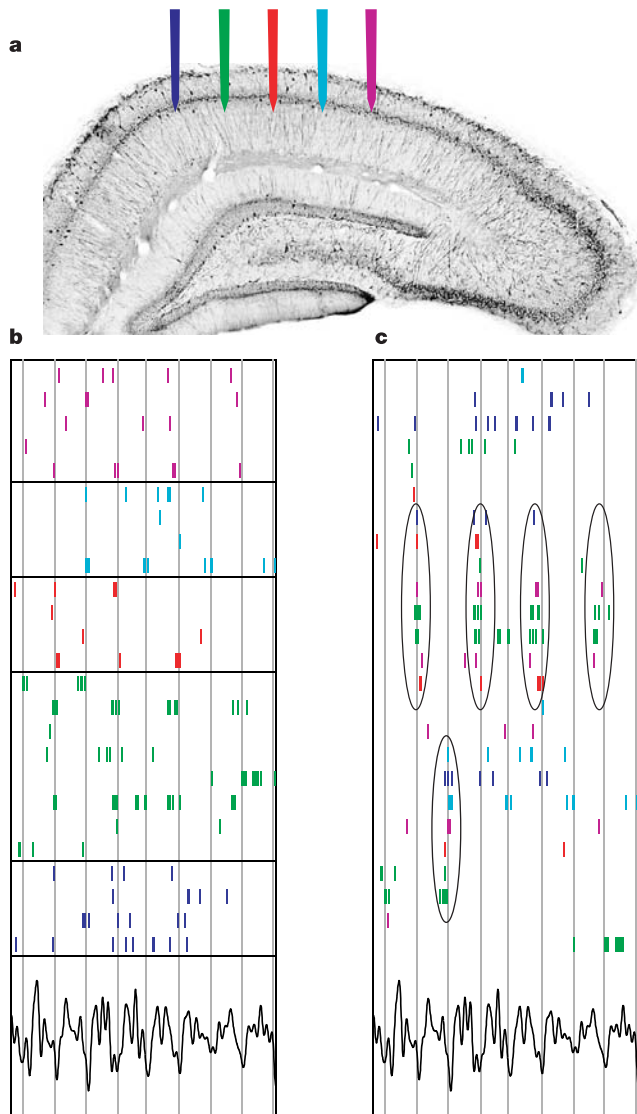


Figure 1 Cell assembly activity in a population. **a**, Location of the recording electrodes. **b**, **c**, Raster plots of 25 pyramidal cells that were active during a 1-s period of spatial exploration out of 68 simultaneously recorded neurons. **b**, Neurons are arranged in order of physical position in the CA1 pyramidal layer (colour-code refers to locations in **a**). Vertical lines indicate troughs of theta waves (bottom trace). Location-specific synchrony is not apparent in the population activity. **c**, The same spike rasters shown in **b**, reordered by stochastic search over all possible orderings to highlight synchrony between anatomically distributed populations. 'Cell assembly' organization is now visible, with repeatedly synchronous firing of some subpopulations (circled).

trains from position by incorporating a position-dependent theta modulation, determined from the neuron's theta phase field⁷. In the first scheme, this is the best prediction that can be made for a cell's spike times. In the second scheme, this prediction should be further improved by peer activity. Figure 4a, b shows an example where the rat traversed the periphery of the neuron's place field, producing a phase pattern considerably more variable than the stereotyped forward phase advancement seen on linear tracks⁵. This irregular spike train was better predicted by peer activity than by the refined spatial prediction (Fig. 4a).

At the group level, to ensure that phase prediction was not compromised by poor theta detection, only cases with high theta signal-to-noise ratio were used (122 cells, 11 recording sessions; Methods). Incorporation of theta phase improved over prediction from the place field alone, as expected ($n = 89$, 73% of cells); however, peer prediction further improved on the refined spatial prediction in 89% of these cells ($n = 79$, Fig. 4c), indicating that neurons show coordinated activity beyond that predicted by their simultaneous phase precession.

Hippocampal population firing patterns are therefore different from those predicted by independent coding of spatial position, even when position-dependent phase modulation is taken into account. This interdependence suggests that the temporal structure of neuronal activity can be understood only at a population level²⁶. Here, a picture emerges of neurons organized into temporary coalitions, whose activity lasts for roughly 25 ms, and whose timing and membership are determined only partially by the animal's location in space. Although a given neuron may be part of several assemblies, there are statistical preferences in partner selection, as reflected by the positively and negatively weighted cells. Could the observed assembly activity result from the precisely timed activation of hippocampal cells by sensory, but non-spatial, stimuli²⁷? For the timing of non-spatial sensory events to be the sole cause of the observed synchronization, the external events would themselves need to show exquisite temporal coordination, on the scale of tens of milliseconds, with respect to the theta rhythm. Rather, we propose that the assembly organization arises from the internal dynamics of neuronal circuits and reflects the operation of non-sensory cognitive phenomena.

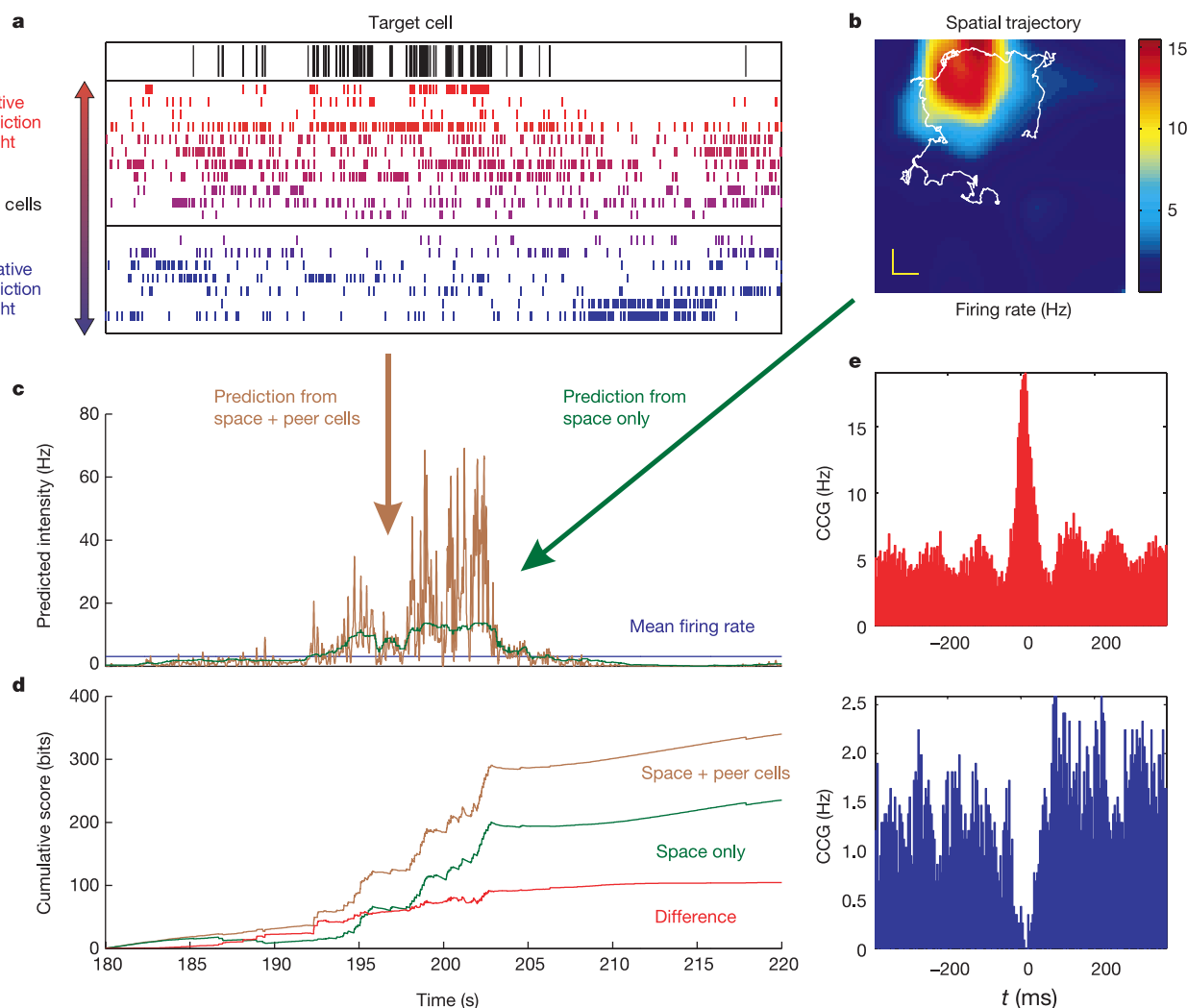


Figure 2 Spike train prediction method. **a**, Activity of the target cell (top) and a group of simultaneously recorded cells (bottom). Each peer cell is assigned a prediction weight by the cross-validation procedure (Methods). The activity of positively or negatively weighted cells predicts increased or decreased the probability of synchronous spikes in the target cell. **b**, The target cell's place field and the animal's trajectory (white trace).

Scale bar, 10 cm. **c**, Instantaneous firing intensity of the target cell predicted from position (green), or position and peer activity (brown). **d**, Prediction quality is quantified by assessing the fit of the actually observed spike train against the prediction. **e**, CCG of the target cell with a positively (top) and a negatively (bottom) predicting cell, showing a peak or trough in the CCG, respectively.

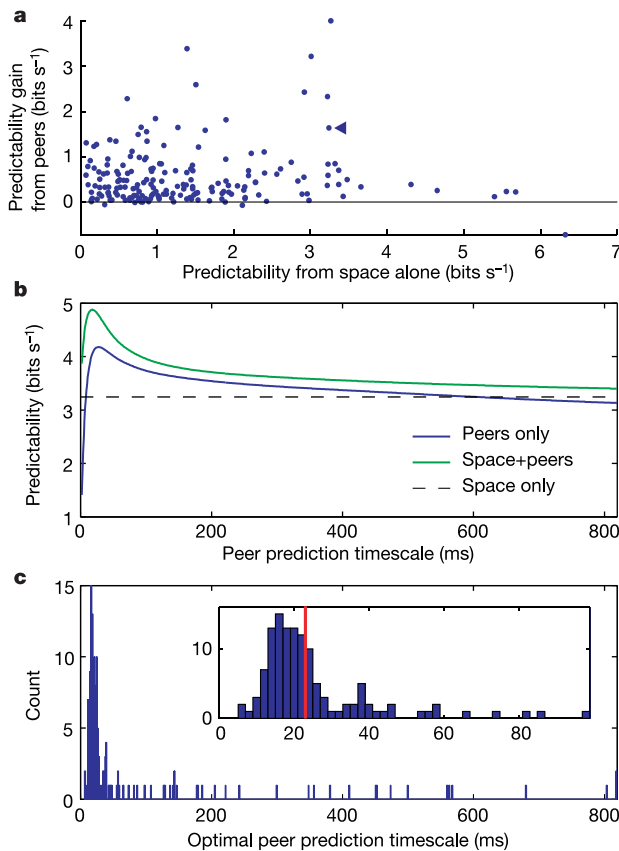


Figure 3 Spike timing is predictable from peer activity. **a**, Gain in predictability of spike trains from peer activity in excess of predictability from location, for all pyramidal cells. Peer activity improved spike train prediction for 98% of cells. **b**, Predictability versus peer prediction timescale for a representative cell (indicated by arrowhead in **a**). Predictability peaks at ~25 ms, indicating that this is the timescale of synchronization for the assemblies in which this cell participates. **c**, Distribution of timescales at which peer activity best improved spike time prediction, for all cells (expanded in inset). The median optimal timescale is 23 ms (red line).

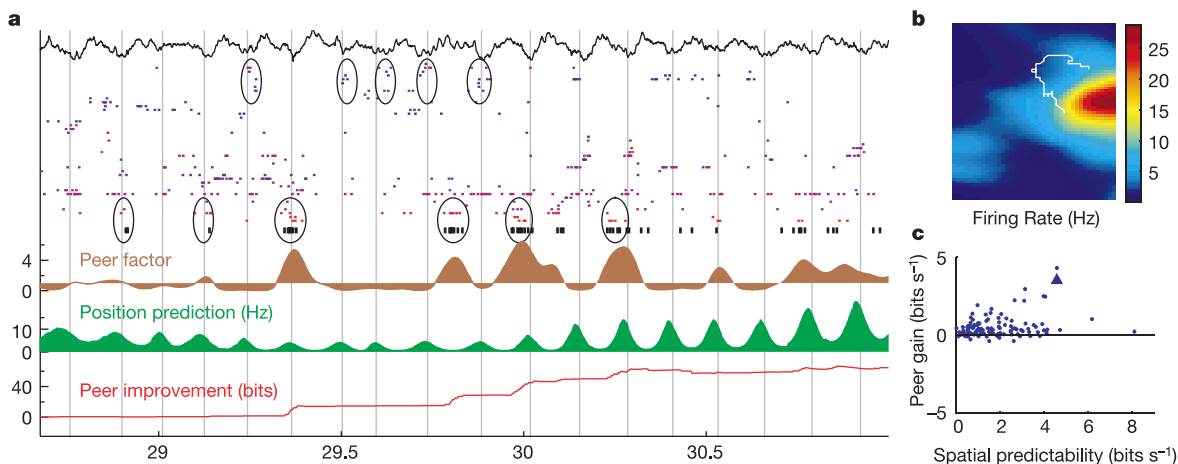


Figure 4 Assembly structure is not fully accounted for by spatially dependent phase modulation. **a**, Top, theta oscillation and spike rasters for a 2-s data segment. Neurons are colour-coded by prediction weight onto the target cell (bottom, black raster; indicated by arrowhead in **c**). Brown indicates peer prediction factor (fill level 1, neutral prediction). Green indicates refined spatial prediction, taking into account theta-phase modulation (fill

The observed 25-ms timescale of synchrony may be physiologically significant. Because it closely matches the membrane time constant of pyramidal neurons in the hippocampal region⁹, activity synchronized with this timescale may be optimal for inducing spiking in downstream neurons. In addition, this timescale matches the period of the hippocampal gamma oscillation¹⁰ and the effective window for synaptic plasticity¹¹. We therefore suggest that assembly activity may be optimal for propagating and storing information in neuronal circuits. □

Methods

Animals and recordings

Male rats (Sprague Dawley or Long Evans) were implanted with either a 64-site silicon probe or several movable tetrodes. All experimental procedures were in accordance with Rutgers University guidelines. Up to 68 cells were recorded simultaneously in 30 recording sessions from six rats, which were either collecting food pellets in an open environment or walking on an elevated square track for food reward. An LED was attached to the head stage to track the position of the animal. Extracellular spikes were extracted from the traces as described^{24,28}. To avoid spurious synchrony, unit activity was not predicted from cells on the same tetrode²⁹. For silicon probes, units were not predicted from peer cells for which the two channels of largest amplitude were the same as for the target cell. We predicted spike trains only for pyramidal cells that had sufficient isolation quality (isolation distance ≥ 20)³⁰ and at least five simultaneously recorded peers. Theta phase was determined by a Hilbert transform⁷ for epochs where the theta-delta power ratio exceeded 6 (ref. 24). Theta was filtered at 8–20 Hz.

Analyses

Analyses were done with custom-written MATLAB and C++ software. Prediction quality was evaluated by tenfold cross-validation. We computed firing rate maps $f(x)$ (place fields) and maps of preferred theta phase and modulation depth (phase fields) from the training set by a local smoothing method^{7,30}. Instantaneous firing intensity was predicted in time bins of size 3.2 ms as the product of a spatial term $f(x_t)$ and a peer prediction term:

$$g\left(\sum_{\alpha} s_{t\alpha} w_{\alpha}\right)$$

where $s_{t\alpha}$ is the temporally smoothed spike count of cell α in time bin t , w_{α} is the weight of cell α , and $g(\eta)$ is $\exp(\eta)$ for $\eta < 0$ and $1 + \eta$ for $\eta \geq 0$. Weights were fit by maximum likelihood on the training set, penalized by the term:

$$-\sum_{\alpha} w_{\alpha}^2 / 4$$

Prediction quality was quantified by the ratio of log-likelihoods:

$$L_f = - \int f(t) dt + \sum_s \log f(t_s)$$

level 0). Red indicates cumulative gain in predictability of target cell spike train using peer activity. **b**, Animal's trajectory for this data segment, superimposed on target cell place field. **c**, Gain in predictability of spike trains by using peer activity for the cell population, plotted against predictability from the refined (phase-modulated) spatial prediction. Use of peer activity further improved the refined spatial prediction.

of the test set spike train under the predicted firing probability relative to a constant probability given by the mean firing rate on the training set. Computed to base 2, this ratio estimates the number of bits a communicator of the test set spike train can save by knowing the predictor variables (position and/or peer cell activity), as compared with knowing only mean rate. Spike rasters were re-ordered for display by stochastic search over all possible orderings, to maximize the fraction of spike pairs between neighbouring rasters that lay within 25 ms (Figs 1b and 4a).

See Supplementary Information for a full description of methods.

Received 14 April; accepted 9 June 2003; doi:10.1038/nature01834.

1. Mainen, Z. F. & Sejnowski, T. J. Reliability of spike timing in neocortical neurons. *Science* **268**, 1503–1506 (1995).
2. Hebb, D. O. *The Organization of Behavior* (Wiley, New York, 1949).
3. Freiwald, W. A., Kreiter, A. K. & Singer, W. Synchronization and assembly formation in the visual cortex. *Prog. Brain Res.* **130**, 111–140 (2001).
4. Engel, A. K., Fries, P. & Singer, W. Dynamic predictions: oscillations and synchrony in top-down processing. *Nature Rev. Neurosci.* **2**, 704–716 (2001).
5. O'Keefe, J. & Recce, M. L. Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus* **3**, 317–330 (1993).
6. Skaggs, W. E., McNaughton, B. L., Wilson, M. A. & Barnes, C. A. Theta phase precession in hippocampal neuronal populations and the compression of temporal sequences. *Hippocampus* **6**, 149–172 (1996).
7. Harris, K. D. *et al.* Spike train dynamics predicts theta-related phase precession in hippocampal pyramidal cells. *Nature* **417**, 738–741 (2002).
8. Mehta, M. R., Lee, A. K. & Wilson, M. A. Role of experience and oscillations in transforming a rate code into a temporal code. *Nature* **417**, 741–746 (2002).
9. Spruston, N. & Johnston, D. Perforated patch-clamp analysis of the passive membrane properties of three classes of hippocampal neurons. *J. Neurophysiol.* **67**, 508–529 (1992).
10. Csicsvari, J., Jamieson, B., Wise, K. D. & Buzsáki, G. Mechanisms of gamma oscillations in the hippocampus of the behaving rat. *Neuron* **37**, 311–322 (2003).
11. Magee, J. C. & Johnston, D. A synaptically controlled, associative signal for Hebbian plasticity in hippocampal neurons. *Science* **275**, 209–213 (1997).
12. Reinagel, P. & Reid, R. C. Precise firing events are conserved across neurons. *J. Neurosci.* **22**, 6837–6841 (2002).
13. Kara, P., Reinagel, P. & Reid, R. C. Low response variability in simultaneously recorded retinal, thalamic, and cortical neurons. *Neuron* **27**, 635–646 (2000).
14. Buracas, G. T., Zador, A. M., DeWeese, M. R. & Albright, T. D. Efficient discrimination of temporal patterns by motion-sensitive neurons in primate visual cortex. *Neuron* **20**, 959–969 (1998).
15. Fenton, A. A. & Muller, R. U. Place cell discharge is extremely variable during individual passes of the rat through the firing field. *Proc. Natl Acad. Sci. USA* **95**, 3182–3187 (1998).
16. Shadlen, M. N. & Newsome, W. T. Noise, neural codes and cortical organization. *Curr. Opin. Neurobiol.* **4**, 569–579 (1994).
17. Lisman, J. E. & Idiart, M. A. Storage of 7 ± 2 short-term memories in oscillatory subcycles. *Science* **267**, 1512–1515 (1995).
18. Redish, A. D. *et al.* Independence of firing correlates of anatomically proximate hippocampal pyramidal cells. *J. Neurosci.* **21**, RC134 [online] (<http://www.jneurosci.org/cgi/content/full/21/10/RC134>) (2001).
19. Hirase, H., Leinekugel, X., Csicsvari, J., Czurko, A. & Buzsáki, G. Behavior-dependent states of the hippocampal network affect functional clustering of neurons. *J. Neurosci.* **21**, RC145 [online] (<http://www.jneurosci.org/cgi/content/full/21/10/RC145>) (2001).
20. O'Keefe, J. & Nadel, L. *The Hippocampus as a Cognitive Map* (Clarendon, Oxford, 1978).
21. Wilson, M. A. & McNaughton, B. L. Dynamics of the hippocampal ensemble code for space. *Science* **261**, 1055–1058 (1993).
22. Brown, E. N., Frank, L. M., Tang, D., Quirk, M. C. & Wilson, M. A. A statistical paradigm for neural spike train decoding applied to prediction from ensemble firing patterns of rat hippocampal place cells. *J. Neurosci.* **18**, 7411–7425 (1998).
23. Cox, D. R. & Isham, V. *Point Processes* (Chapman and Hall, London, 1980).
24. Csicsvari, J., Hirase, H., Czurko, A., Mamia, A. & Buzsáki, G. Oscillatory coupling of hippocampal pyramidal cells and interneurons in the behaving rat. *J. Neurosci.* **19**, 274–287 (1999).
25. Jensen, O. & Lisman, J. E. Position reconstruction from an ensemble of hippocampal place cells: contribution of theta phase coding. *J. Neurophysiol.* **83**, 2602–2609 (2000).
26. deCharms, R. C. & Zador, A. Neural representation and the cortical code. *Annu. Rev. Neurosci.* **23**, 613–647 (2000).
27. Wood, E. R., Dudchenko, P. A. & Eichenbaum, H. The global record of memory in hippocampal neuronal activity. *Nature* **397**, 613–616 (1999).
28. Harris, K. D., Henze, D. A., Csicsvari, J., Hirase, H. & Buzsáki, G. Accuracy of tetrode spike separation as determined by simultaneous intracellular and extracellular measurements. *J. Neurophysiol.* **84**, 401–414 (2000).
29. Quirk, M. C. & Wilson, M. A. Interaction between spike waveform classification and temporal sequence detection. *J. Neurosci. Methods* **94**, 41–52 (1999).
30. Harris, K. D., Hirase, H., Leinekugel, X., Henze, D. A. & Buzsáki, G. Temporal interaction between single spikes and complex spike bursts in hippocampal pyramidal cells. *Neuron* **32**, 141–149 (2001).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank J. E. Lisman, D. L. Buhl, S. M. Montgomery, P. E. Bartho and I. Creese for comments on the manuscript. This work was supported by grants from the National Institutes of Health

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to G.B. (buzsaki@axon.rutgers.edu).

Role of the prolyl isomerase Pin1 in protecting against age-dependent neurodegeneration

Yih-Cherng Liou^{1*†}, Anyang Sun^{1,2*}, Akihide Ryo^{1†}, Xiao Zhen Zhou¹, Zhao-Xue Yu³, Han-Kuei Huang⁴, Takafumi Uchida⁵, Roderick Bronson⁶, Guoying Bing², Xiaojiang Li³, Tony Hunter⁴ & Kun Ping Lu¹

¹Cancer Biology Program, Department of Medicine, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts 02215, USA

²Department of Anatomy & Neurobiology, University of Kentucky, Lexington, Kentucky 40536, USA

³Department of Human Genetics, Emory University, Atlanta, Georgia 30322, USA

⁴Molecular and Cell Biology Laboratory, Salk Institute, La Jolla, California 92037, USA

⁵Department of Pathology, Tohoku University, Sendai 980-8575, Japan

⁶Tufts University School of Veterinary Medicine, Boston, Massachusetts 01536, USA

* These authors contributed equally to this work

† Present address: Department of Biochemistry, National University of Singapore, 117597, Singapore (Y.-C.L.); Department of Pathology, Yokohama City University, Kanagawa 236-0004, Japan (A.R.)

The neuropathological hallmarks of Alzheimer's disease and other tauopathies include senile plaques and/or neurofibrillary tangles^{1–4}. Although mouse models have been created by over-expressing specific proteins including β -amyloid precursor protein, presenilin and tau^{1–10}, no model has been generated by gene knockout. Phosphorylation of tau and other proteins on serine or threonine residues preceding proline seems to precede tangle formation and neurodegeneration in Alzheimer's disease^{11–14}. Notably, these phospho(Ser/Thr)-Pro motifs exist in two distinct conformations, whose conversion in some proteins is catalysed by the Pin1 prolyl isomerase^{15–17}. Pin1 activity can directly restore the conformation and function of phosphorylated tau or it can do so indirectly by promoting its dephosphorylation, which suggests that Pin1 is involved in neurodegeneration^{14,18,19}; however, genetic evidence is lacking. Here we show that Pin1 expression is inversely correlated with predicted neuronal vulnerability and actual neurofibrillary degeneration in Alzheimer's disease. Pin1 knockout in mice causes progressive age-dependent neuropathy characterized by motor and behavioural deficits, tau hyperphosphorylation, tau filament formation and neuronal degeneration. Thus, Pin1 is pivotal in protecting against age-dependent neurodegeneration, providing insight into the pathogenesis and treatment of Alzheimer's disease and other tauopathies.

Pin1-catalysed prolyl isomerization can regulate the function and/or dephosphorylation of some phosphoproteins, many of which are also recognized by the mitosis- and phosphorylation-specific monoclonal antibody MPM-2. Notably, induction of MPM-2 epitopes is a prominent common feature of Alzheimer's disease (AD), frontotemporal dementia with Parkinsonism linked to chromosome 17, Down's syndrome, corticobasal degeneration, progressive supranuclear palsy and Pick's disease^{13,14}. In fact, the pattern of tau phosphorylation in AD is similar to that in mitotic cells^{12,14}. Taking these observations together with the reduced amount of soluble Pin1 in brains at a late stage of AD¹⁸, we previously proposed that Pin1 might protect against neurodegeneration^{14,18}. However, it has been reported that in AD hippocampal expression of Pin1 occurs primarily in a few tangle-free degenerative neurons in CA1 and CA2, and not in CA3 and CA4 non-degenerative neurons, and it has been proposed that Pin1 promotes neurodegeneration²⁰. Thus, the neuronal function of Pin1 remains elusive.

Neurons in different subregions of the hippocampus and neo-

cortex are known to have differential vulnerability to neurofibrillary degeneration in AD²¹. To examine the relationship between this predicted vulnerability and Pin1 expression, we examined the expression of Pin1 in normal human hippocampus and parietal cortex by immunostaining. Pin1 was detected in the cytoplasm and the nucleus of neurons (ref. 18 and Fig. 1). Notably, Pin1 expression showed an obvious subregional difference in all sections from eight normal brains (Fig. 1a and Supplementary Figs S1a and S2). In the hippocampus, expression of Pin1 was relatively higher in CA4, CA3, CA2 and presubiculum, and lower in CA1 and subiculum. In the parietal cortex, expression of Pin1 was relatively higher in layer IIIb-c neurons, and lower in layer V neurons (Supplementary Fig. S3a). The subregions with low expression of Pin1 are known to be prone to neurofibrillary degeneration in AD, whereas those containing high Pin1 expression are spared, suggesting that there is an inverse correlation between Pin1 expression and predicted vulnerability.

To extend this correlation, we immunostained ten AD-affected brain sections with both antibodies against Pin1 and a phospho-Tau

antibody AT8 that detects early neurofibrillary degeneration. Tangle-bearing neurons were enriched in CA1 and subiculum of the hippocampus and in layer V of the parietal cortex (Fig. 1b and Supplementary Figs S1b and S3b). We also observed very high Pin1 immunoreactivity in cytoplasmic granules in a few neurons, mainly in CA1 (data not shown), which is probably due to the sequestration of Pin1 by strong MPM-2 epitopes in these granules²⁰.

In contrast to previous data²⁰, however, we found that Pin1 expression was readily detected in CA3 and CA4 subregions and, in fact, was higher in these regions than in tangle-rich subregions (Fig. 1b and Supplementary Fig. S1b). Overall, 96% of pyramidal neurons that contained relatively more Pin1 lacked tangles, whereas 71% of neurons that contained relatively less Pin1 had tangles (Fig. 1c). Even in the tangle-prone CA1 and subiculum of the hippocampus, expression of Pin1 was still lower in most tangle-bearing neurons than in tangle-free neurons (Fig. 1d). Expression of Pin1 was also relatively higher in tangle-sparing layer IIIb-c neurons, and lower in tangle-rich layer V neurons in parietal cortex (Supplementary Fig. S3b). These results suggest that there is an inverse correlation between Pin1 expression and actual neurofibrillary degeneration in AD.

The above results suggest that Pin1 might protect against neurodegeneration. To test this idea, we examined the neuronal phenotypes of *Pin1*^{-/-} mice. These mice develop several age-dependent phenotypes, including retinal atrophy²². Because retinal atrophy can

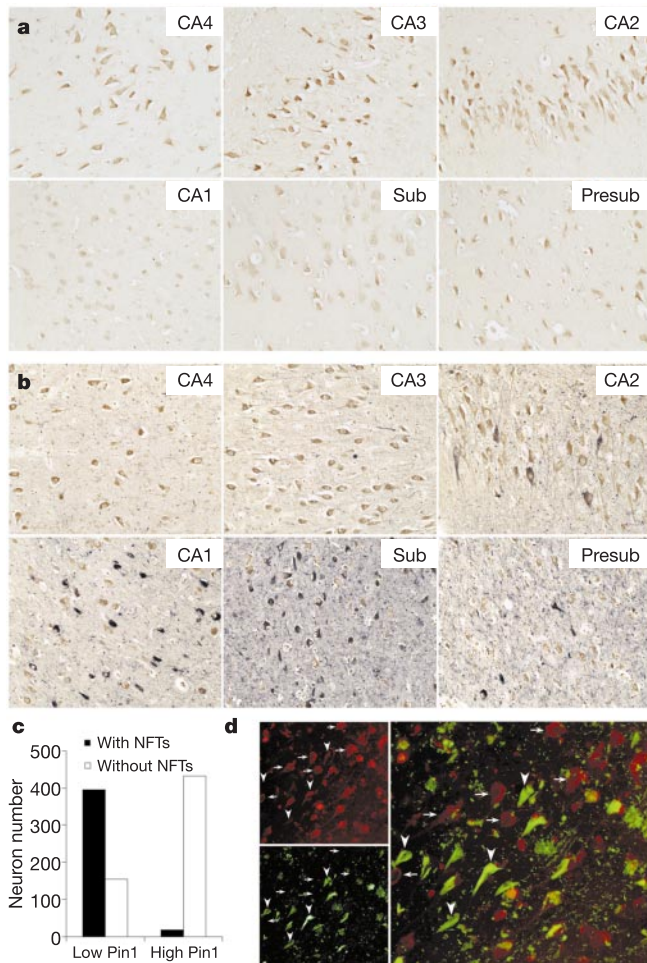


Figure 1 Inverse correlation of Pin1 expression with predicted neuronal vulnerability in normally aged brains and actual neurofibrillary degeneration in AD. **a, b**, Hippocampal sections from normal (**a**) or AD (**b**) brains were immunostained with antibodies against Pin1 (**a**) or antibodies against Pin1 (brown) and AT8 (purple) (**b**). **c**, Relationship between Pin1 immunoreactivity and NFTs in AD hippocampus. About 1,000 pyramidal neurons in AD were randomly selected and evaluated for AT8-positive or AT8-negative neurons, and Pin1-light (low) or Pin1-intense (high) immunoreactivity. **d**, Relationship between Pin1 immunoreactivity and NFTs in tangle-rich CA1 region detected by antibodies against Pin1 (red) and AT8 (green), respectively. Pin1 expression in most tangle-bearing neurons (arrowheads) was lower than that in tangle-free neurons (arrows).

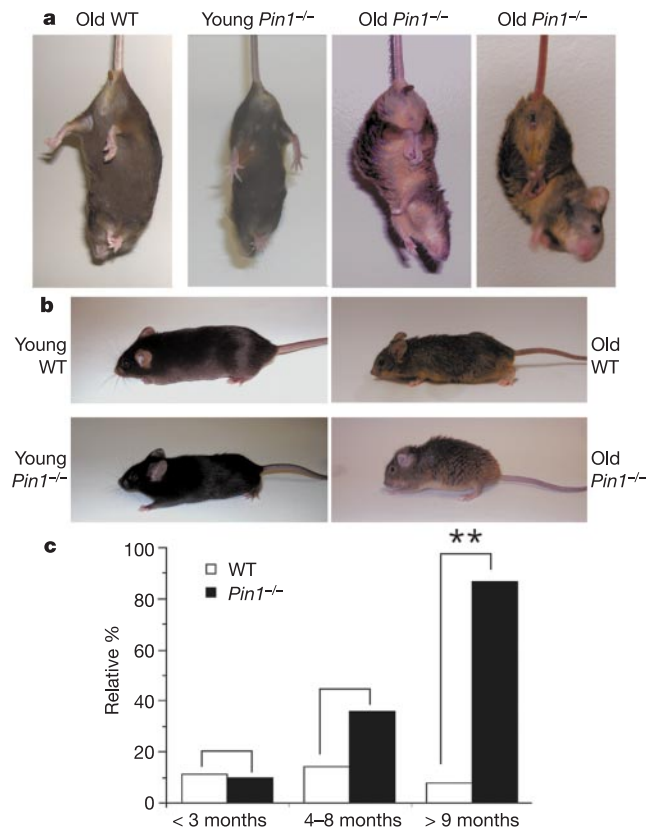


Figure 2 Age-dependent motor and behavioural deficits in *Pin1*^{-/-} mice. **a**, Abnormal limb-clasping reflexes in old *Pin1*^{-/-} mice. When lifted by the tail, young wild-type and *Pin1*^{-/-} mice (2–3 months) and old wild-type mice (9–14 months) acted normally by extending their hind limbs and body, but old *Pin1*^{-/-} mice flexed their legs to the trunk or tightened the back limbs to their bodies. **b**, Hunched postures shown by old *Pin1*^{-/-} mice, but not by young *Pin1*^{-/-} or any wild-type mice. **c**, Age-dependent motor disturbance. More than ten mice at different ages were placed on a hanging bar, and the percentage of the mice that fell off during the 1-min test period was recorded. Double asterisk, $P < 0.01$.

be a feature of neurodegeneration, we thought that *Pin1*^{-/-} mice might show other neuronal phenotypes. Indeed, *Pin1*^{-/-} mice, but not their wild-type littermates, showed progressive age-dependent motor and behavioural deficits, which included abnormal limb-clasping reflexes (Fig. 2a), hunched postures (Fig. 2b), reduced mobility and eye irritation. When subjected to a hang test⁶, most old, but not young, *Pin1*^{-/-} mice fell after grasping the rope only briefly (Fig. 2c). Thus, *Pin1*^{-/-} mice develop progressive age-dependent motor and behavioural deficits, as do tau transgenic mice^{6,10}.

We examined whether *Pin1*^{-/-} mice show age-dependent neuronal loss. The number of NeuN (neuron-specific nuclear protein)-positive neurons was significantly decreased in the parietal cortex of old, but not young, *Pin1*^{-/-} mice (Fig. 3a, b). A similar neuronal degeneration was found in spinal cords of *Pin1*^{-/-} mice (Supplementary Fig. S4a, b). Nissl staining also confirmed neuronal degeneration, as shown by the abnormally stained cytoarchitecture of *Pin1*^{-/-} neurons (Supplementary Fig. S4c, arrows). Some *Pin1*^{-/-} neurons had swollen cell bodies (data not shown), as observed in tau transgenic mice¹⁰; however, no obvious neuronal loss was found in other brain regions, notably cerebellum (data not shown), although there was *Pin1* expression in wild-type mice (Supplementary Fig. S5).

We used electron microscopy to confirm degeneration. In contrast to wild-type neurons (Fig. 3c), many *Pin1*^{-/-} neurons showed dark, degenerating granules or organelles adjacent to the nucleus (Fig. 3d) and often had autophagic vacuoles (Fig. 3e), consistent with degenerated lysosomes. Degeneration was also observed in

some axons (Fig. 3f). Another notable change found mostly in *Pin1*^{-/-} neuronal processes, but not in wild-type controls, was the presence of electron-dense structures comprising compact and radiating filament-like structures without other visible organelles or vesicles (Fig. 3g, h). These ultrastructural changes confirm that there is degeneration in *Pin1*^{-/-} neurons.

The findings that *Pin1*^{-/-} mice develop neuronal degeneration in an age-dependent manner in some brain regions suggest that the effects are specific. The ability of Pin1 to promote dephosphorylation of MPM-2 epitopes¹⁹ suggests that the deletion of *Pin1* might lead to an accumulation of MPM-2 epitopes, which is an early common characteristic of AD and related disorders^{12,14,23}. Indeed, total MPM-2 reactivity was about threefold higher in *Pin1*^{-/-} brain lysates than in control brain lysates (Fig. 4a), a result that was confirmed by immunostaining (Fig. 4g). These results indicate that *Pin1* knockout leads to neuronal induction of MPM-2 epitopes.

To examine the mechanisms underlying neurodegeneration, we focused on tau for the following reasons. First, tau-related pathologies are a hallmark of AD and other tauopathies^{2,3} and have been characterized in mice⁵⁻¹⁰; notably, the phenotypes of transgenic tau mice are similar to those of *Pin1*^{-/-} mice. Second, tau is a principal MPM-2 antigen¹³ and a well-characterized substrate of Pin1 (refs 18, 19). Pin1 specifically acts on the phospho(Thr 231)-Pro motif in tau and induces a conformational change, thereby restoring tau function and promoting its dephosphorylation through the conformational specificity of phosphatases such as PP2A^{18,19}. Reducing PP2A activity also increases phosphorylation in tau in mice²⁴. Thus, tau might be aberrantly phosphorylated and show abnormal conformations in *Pin1*^{-/-} mice. To examine this possibility, we isolated sarcosyl-insoluble extracts from brains of *Pin1*^{-/-} and wild-type mice and analysed them by immunoblotting (Fig. 4b-f).

All sarcosyl-insoluble tau isoforms from *Pin1*^{-/-} mice had much slower mobility on SDS gels than did wild-type controls (Fig. 4b), suggesting an increase in the total phosphorylation of tau. This was confirmed by phosphatase treatment and immunoblotting with phospho-specific monoclonal antibodies against tau (Fig. 4c, d). AT8 and AT180 strongly recognized the slower migrating tau isoforms, whereas TG3 selectively recognized the slowest migrating tau species (relative molecular mass ~68,000; *M_r* ≈ 68K), which was also recognized by PHF-1 (Fig. 4d). The TG3 epitope was also induced in an age-dependent manner, and phosphatase treatment significantly, although not completely, reduced the TG3 signal (Fig. 4e, f). Finally, the 68K form of tau in *Pin1*^{-/-} brains was also strongly recognized by Alz50 and MC1 (Fig. 4d), which detect neurofibrillary tangle (NFT)-specific conformations^{25,26}.

These results indicate that the 68K form of tau in *Pin1*^{-/-} brain is hyperphosphorylated and contains NFT conformations. Of note, tau in NFTs of AD is notoriously resistant to complete dephosphorylation²⁷ and a similar 68K species (A68) in human AD is the defining component of paired helical filaments (PHFs) and is also recognized by Alz50 (refs 25, 26). Immunostaining showed strong immunoreactivity towards AT180, AT8, MC1 and Alz50 in the somatodendritic region of *Pin1*^{-/-}, but not wild-type, neurons in the parietal cortex, hippocampus, brainstem and spinal cord (Fig. 4g). The presence of phospho-Tau in the cytoplasm and axons of *Pin1*^{-/-}, but not wild-type, neurons was confirmed by immunogold electron microscopy (Fig. 3c, d and Supplementary Fig. S6). These results indicate that *Pin1* knockout causes age-dependent tau hyperphosphorylation and NFT-specific conformations.

To explore the mechanisms that induce MPM-2 and tau phosphoepitopes, we compared protein kinase and phosphatase activity in *Pin1*^{-/-} and wild-type brains. Although there was no significant increase in the activity of total kinases, cyclin-dependent kinases (CDKs) or glycogen synthase kinase 3β (GSK-3β; Supplementary Fig. S7a, b), there was a significant decrease in phosphatase activity towards phospho(Ser/Thr)-Pro motifs in tau, but

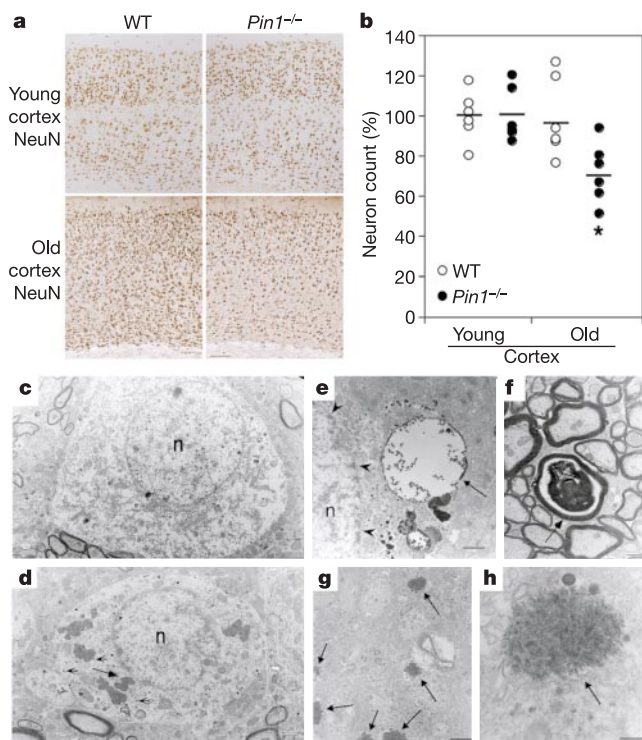


Figure 3 Age-dependent neuronal degeneration in *Pin1*^{-/-} mice. **a, b**, Matched parietal cortex from wild-type and *Pin1*^{-/-} mice was stained for NeuN (**a**) and neurons were counted (**b**). Asterisk, *P* < 0.01. **c–h**, Ultrastructure and AT8 immunogold labelling of degenerative neurons in *Pin1*^{-/-} hippocampus. **c**, Wild-type neuron not labelled by AT8. **d**, *Pin1*^{-/-} neuron labelled with AT8 immunogold (short arrows) and containing dark and degenerated organelles (long arrows) in the cytoplasm and near the nucleus (n). **e**, Autophagic vacuole (arrow) near the nuclear membrane (arrowheads). **f**, Axon degeneration (arrow). **g, h**, Neuron containing several compact and radiating structures (arrows, **g**) composed of radiating filament-like structures (arrow, **h**). Scale bars, 100 μm (**a**); 0.6 μm (**c, d**); 1.2 μm (**e, f**); 2 μm (**g**); 355 nm (**h**).

not towards a non-phospho(Ser/Thr)-Pro phosphopeptide, in *Pin1*^{-/-} brain lysates before obvious neurodegeneration (Fig. 4h). These results indicate that Pin1 specifically affects the dephosphorylation of phospho(Ser/Thr)-Pro motifs, consistent with the idea that Pin1 is required for efficient dephosphorylation of tau¹⁹.

The findings that *Pin1*^{-/-} neurons are strongly immunoreactive towards phospho- or NFT-specific monoclonal antibodies against tau suggests that they might contain tau filaments. To examine this possibility, we used Gallyas silver staining and thioflavin-S staining

to detect NFTs^{6,9,29}. In *Pin1*^{-/-} mice (Fig. 5b–d), but not in wild-type controls (Fig. 5a), we observed some Gallyas-positive neurons in the hippocampus, thalamus and brainstem. In addition, a few *Pin1*^{-/-}, but not wild-type, neurons in the hippocampus, spinal cord and especially entorhinal cortex (ERC) were also positive for thioflavin-S staining (Fig. 5h); this positive staining was observed in four out of seven old *Pin1*^{-/-} mice, but not in six wild-type mice examined. Notably, the pattern of thioflavin-S staining was similar to that observed in tau transgenic mice¹⁰.

These results show that *Pin1*^{-/-} neurons have NFT-like pathologies. Our attempt to visualize NFTs by *in situ* electron microscopy failed, possibly because the number of NFTs and/or neurons that developed late stages of NFTs was low, as indicated by thioflavin-S staining (Fig. 5e–h). To confirm the formation of tau filaments, we subjected sarcosyl-insoluble tau isolated from brain extracts to electron microscopy and immunogold electron microscopy. In extracts from 8 out of 12 old *Pin1*^{-/-} brains, but not wild-type brains, we readily found filaments that were twisted or straight and about 15-nm wide (Fig. 5i, j). Further immunogold electron microscopy showed that the filaments were labelled with AT8 and AT180 (Fig. 5k, l), confirming that they were tau filaments. These results show that loss of Pin1 function causes the formation of endogenous tau filaments in mice.

In summary, we have shown that Pin1 expression inversely correlates with neuronal vulnerability in normal brain, and also with neurofibrillary degeneration in AD-affected brain. In addition, *Pin1*^{-/-} mice develop age-dependent neuropathy, which is characterized clinically by motor and behavioural deficits and pathologically by tau hyperphosphorylation, tau filament formation and

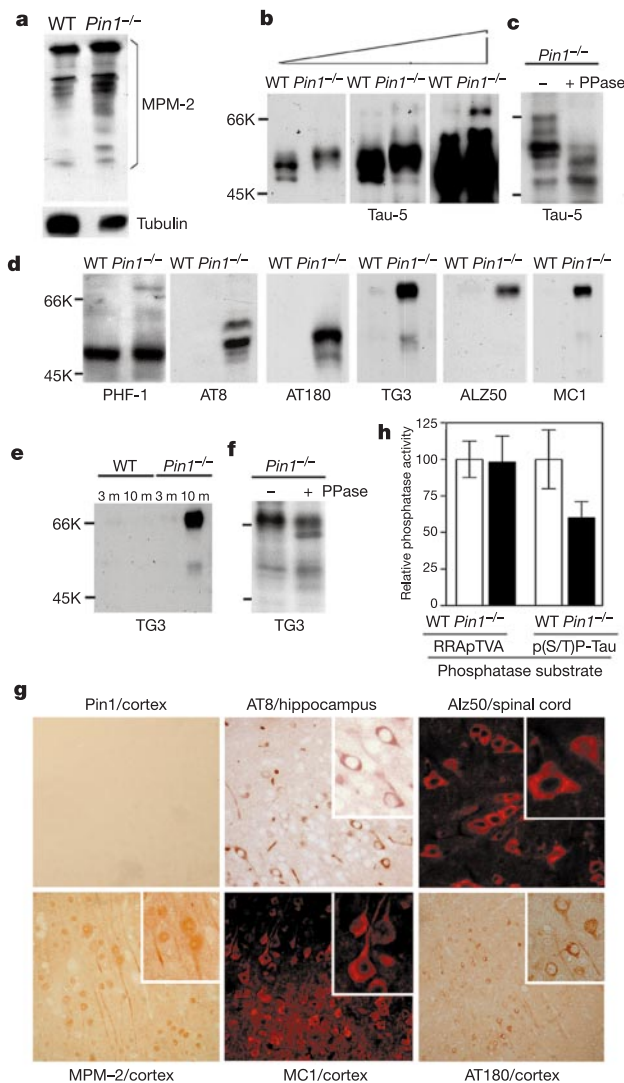


Figure 4 MPM-2 induction, tau hyperphosphorylation, NFT-specific conformations and reduced phosphatase activity toward the phospho(Ser/Thr)-Pro motif in *Pin1*^{-/-} brain. **a**, Soluble brain extracts from age-matched old wild-type and *Pin1*^{-/-} mice were analysed by immunoblotting with MPM-2. **b**, **d**, Sarcosyl-insoluble fractions were prepared from age-matched wild-type and *Pin1*^{-/-} brains and analysed by immunoblotting with Tau-5 monoclonal antibodies against total tau (**b**) or various phosphorylation- and/or NFT-specific monoclonal antibodies against tau (**d**). **c**, **f**, Sarcosyl-insoluble tau was pretreated with phosphatases before being analysed by immunoblotting. **e**, Age-dependent induction of the TG3 epitope. **g**, Subcellular localization of MPM-2 epitopes, tau phosphoepitopes and NFT-conformation epitopes in *Pin1*^{-/-} neurons, as determined by immunostaining. There was no positive staining in wild-type neurons (not shown). **h**, Activity towards phospho(Ser/Thr)-Pro motifs, but not towards a non-phospho(Ser/Thr)-Pro motif, is lower in brain lysates from 7-month-old *Pin1*^{-/-} mice than in those from wild-type controls, as assayed with the indicated substrates.

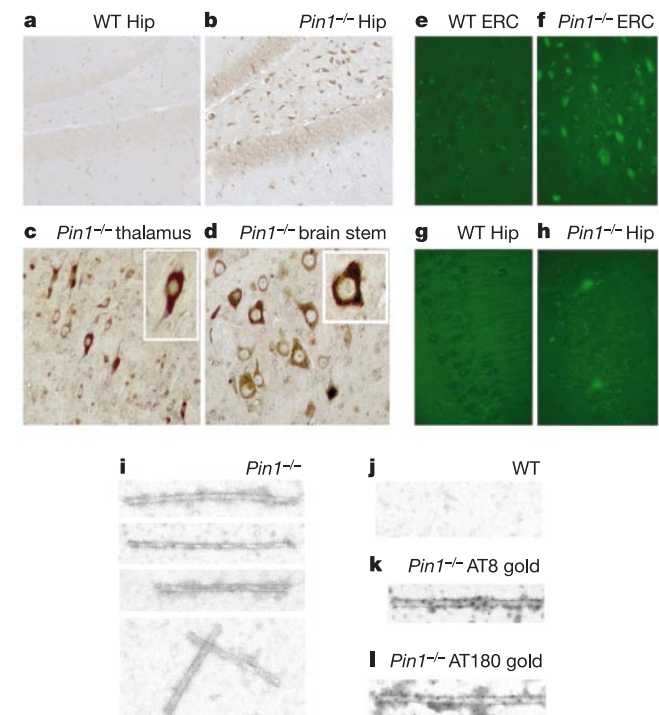


Figure 5 NFT-like pathologies and tau filaments in *Pin1*^{-/-} neurons. **a–d**, Positive Gallyas silver staining of *Pin1*^{-/-} neurons. Different regions of old wild-type (**a**) or *Pin1*^{-/-} (**b–d**) brains were subjected to Gallyas silver staining. **e–h**, Positive thioflavin-S staining of *Pin1*^{-/-} neurons. Different regions of old wild-type (**e**, **g**) or *Pin1*^{-/-} (**f**, **h**) brains were subjected to thioflavin-S staining. **i–j**, Tau filaments isolated from *Pin1*^{-/-} brains. Sarcosyl-insoluble extracts were prepared from old *Pin1*^{-/-} (**i**) and wild-type (**j**) mice and examined by electron microscopy. **k**, **l**, Phosphorylated tau in the filaments. Sarcosyl-insoluble extracts were subjected to immunogold staining with AT8 (**k**) or AT180 (**l**) and then examined by electron microscopy. ERC, entorhinal cortex; Hip, hippocampus.

neuronal degeneration in brain and spinal cord. Notably, many of these neuronal phenotypes are highly similar to those induced by transgenic overexpression of tau or its mutants^{5–10}. These results provide genetic evidence for the critical role of Pin1 in protecting against age-dependent neurodegeneration. To our knowledge, this is the first clear demonstration that endogenous mouse tau can form tau filaments. Thus, Pin1 is the first protein whose deletion has been shown to cause age-dependent neurodegeneration and tau pathologies. In addition, our finding that Pin1-mediated post-phosphorylation regulation is pivotal in maintaining normal neuronal function underscores the role of protein phosphorylation in neurodegenerative diseases and offers insight into the pathogenesis and treatment of AD and other tauopathies.

Given the phenotypic similarity between *Pin1*^{−/−} and tau transgenic mice, tau-related pathologies are probably important in *Pin1*^{−/−}-induced neurodegeneration, although other Pin1-related deficits might also contribute to this process²². Manipulation of either tau kinases or phosphatases has been shown to increase phospho-Tau in mice^{14,24,29,30}. We have now shown that deletion of *Pin1* reduces phosphatase activity specifically towards phospho(-Ser/Thr)-Pro motifs and induces tau hyperphosphorylation, NFT conformations and tau filament formation. These results suggest that tau is normally regulated by dynamic phosphorylation and dephosphorylation. If Pin1 function is low, as in the CA1 region of AD-affected brains, or absent, as in *Pin1*^{−/−} mice, some phospho(-Ser/Thr)-Pro motifs in phospho-Tau might not be isomerized and thus exist in aberrant conformations. As a result, phospho-Tau would not be properly dephosphorylated and/or functionally restored, leading to tau hyperphosphorylation, NFT formation and neurodegeneration.

Although there is clear evidence of the age-dependent accumulation of abnormal and hyperphosphorylated tau, tau filaments and neurodegeneration in *Pin1*^{−/−} brains, the patterns are not exactly identical to those observed in human AD or in human tau transgenic models. This might be due to differences in tau isoforms in mice or might reflect differences in the phosphorylation and aggregation of mouse tau. In addition, multiple and interactive factors probably contribute to tau hyperphosphorylation and neurodegeneration in human AD, and these factors might not be present in *Pin1*^{−/−} mice. Further studies, including crossing *Pin1*^{−/−} and *Pin1* transgenic mice with existing mouse models of amyloid- β or tauopathy, should help to elucidate the molecular mechanisms of AD and other tauopathies and might lead to the development of new therapies. □

Methods

Human samples and *Pin1*^{−/−} mouse strains

Paraffin-embedded human samples of the hippocampus and parietal cortex were a gift from the Alzheimer's Disease Research Center (University of Kentucky) and consisted of ten cases of AD and eight controls, with a mean age at death of 78 $\pm$ 9 years and 79 $\pm$ 9 years, respectively. The mean postmortem intervals were 2.8 h for AD and 2.9 h for controls. All AD subjects met the clinical and neuropathological National Institute on Aging-Reagan Institute (NIA-RI) criteria for AD. Control subjects had no evidence of neurological disorders. Our study using human samples has been approved by our Institutional Review Board. The *Pin1*^{−/−} mice are on a mixed 129/Sv and C57L/B6 background^{22,28}. Young and old mice were 2–3 and 9–14 months, respectively. All results were reproduced in several mice.

Immunohistochemistry and immunofluorescent staining

We carried out conventional immunohistochemistry on human samples by using affinity-purified polyclonal and monoclonal antibodies against Pin1, as described¹⁸. For double-labelling experiments, sections were stained first with antibodies against Pin1 (brown), and then with AT8, which was visualized by the nickel-intensifying method (purple). For immunofluorescent double immunostaining, sections were incubated with an affinity-purified antibody against Pin1 and a monoclonal antibody against AT8, which were visualized by Cy3- and fluorescein isothiocyanate (FITC)-conjugated secondary antibodies. For mouse tissues, deparaffinized or floating brain sections were immunostained with various antibodies using an ABC kit (Vector Labs) or Cy3-conjugated secondary antibodies²².

Immunoblot analysis

We prepared sarcosyl-insoluble extracts as described^{7,8,10}. In brief, brain tissues were homogenized in a buffer containing 10 mM Tris-HCl (pH 7.4), 0.8 M NaCl, 1 mM EGTA and 10% sucrose. After centrifugation, the supernatants were added to 1% N-lauroylsarcosinate, and sarcosyl-insoluble extracts were collected by centrifugation and analysed by immunoblotting¹⁸. Alz50, MC1, TG3 and PHF-1 monoclonal antibodies against tau were a gift from P. Davies (Albert Einstein College of Medicine). AT8 and AT180 (Innogenetics) and Tau-5 (Biosource) antibodies were purchased. Tau-5 is specific for both phosphorylated and non-phosphorylated tau, PHF-1 for phosphoSer-396 and phosphoSer-404, AT8 for phosphoSer-199 and phosphoSer-202, AT180 for phosphoThr-231, TG3 for phosphoThr-231 in a NFT-specific conformation, and Alz50 and MC1 for NFT-specific conformations.

Hang test, neuron count and Nissl staining

We carried out the hang test, neuronal counting and Nissl staining as described^{5,6}. For counting neurons, coronal sections at different regions of brains from age-matched *Pin1*^{−/−} and wild-type littermates were processed in parallel for NeuN staining. The inner layer of the parietal cortex (SIBF) at 1.8-mm posterior to Bregma was selected, neurons were counted in comparable areas of each mouse, and an average number of two adjacent fields were obtained for each region of each mouse. For the spinal cord, the numbers of large neurons per anterior horn were counted and an average of two nearby sections was calculated for each mouse.

Kinase and phosphatase assays

We assayed total kinase activity in brain lysates by autophosphorylation in the presence of Mg²⁺ and [γ -³²P]ATP, and assayed CDK and GSK-3 β activity towards tau after purification with p13^{suc1} beads and antibodies against GSK-3 β , respectively, as described¹⁸. Phosphatase activity towards the phosphopeptide RRAPTVTA (Promega; ref. 24) and phosphatase activity towards tau phosphorylated by Cdc2 (refs 18, 19) were assayed as described.

Gallyas silver staining and thioflavin-S staining

Gallyas silver staining⁶ and thioflavin-S staining¹⁰ were done as described.

Electron microscopy

To observe tau filaments, we resuspended sarcosyl-insoluble extracts isolated from *Pin1*^{−/−} and control mouse brains and placed them on carbon-coated grids. The extracts were stained with phosphotungstic acid and observed by electron microscopy and immunogold electron microscopy with AT8 and AT180 (refs 7, 10).

Received 14 March; accepted 28 May 2003; doi:10.1038/nature01832.

- Selkoe, D. J. The cell biology of β -amyloid precursor protein and presenilin in Alzheimer's disease. *Trends Cell Biol.* **8**, 447–453 (1998).
- Mandelkow, E. M. & Mandelkow, E. Tau in Alzheimer's disease. *Trends Cell Biol.* **8**, 425–427 (1998).
- Lee, V. M., Goedert, M. & Trojanowski, J. Q. Neurodegenerative tauopathies. *Annu. Rev. Neurosci.* **24**, 1121–1159 (2001).
- Wong, P. C., Cai, H., Borchelt, D. R. & Price, D. L. Genetically engineered mouse models of neurodegenerative diseases. *Nature Neurosci.* **5**, 633–639 (2002).
- Ishihara, T. et al. Age-dependent emergence and progression of a tauopathy in transgenic mice overexpressing the shortest human tau isoform. *Neuron* **24**, 751–762 (1999).
- Lewis, J. et al. Neurofibrillary tangles, amyotrophy and progressive motor disturbance in mice expressing mutant (P301L) tau protein. *Nature Genet.* **25**, 402–405 (2000).
- Gotz, J., Chen, E., Barmettler, R. & Nitsch, R. M. Tau filament formation in transgenic mice expressing P301L tau. *J. Biol. Chem.* **276**, 529–534 (2001).
- Lewis, J. et al. Enhanced neurofibrillary degeneration in transgenic mice expressing mutant tau and APP. *Science* **293**, 1487–1491 (2001).
- Gotz, J., Chen, E., van Dorpe, J. & Nitsch, R. M. Formation of neurofibrillary tangles in P301L tau transgenic mice induced by A β 42 fibrils. *Science* **293**, 1491–1495 (2001).
- Allen, B. et al. Abundant tau filaments and nonapoptotic neurodegeneration in transgenic mice expressing human P301S tau protein. *J. Neurosci.* **22**, 9340–9351 (2002).
- Bancher, C. et al. Accumulation of abnormally phosphorylated tau precedes the formation of neurofibrillary tangles in Alzheimer's disease. *Brain Res.* **477**, 90–99 (1989).
- Preuss, U. & Mandelkow, E. M. Mitotic phosphorylation of tau protein in neuronal cell lines resembles phosphorylation in Alzheimer's disease. *Eur. J. Cell Biol.* **76**, 176–184 (1998).
- Vincent, I., Zheng, J. H., Dickson, D. W., Kress, Y. & Davies, P. Mitotic phosphopeptides precede paired helical filaments in Alzheimer's disease. *Neurobiol. Aging* **19**, 287–296 (1998).
- Lu, K. P., Liou, Y. C. & Vincent, I. Proline-directed phosphorylation and isomerization in mitotic regulation and in Alzheimer's disease. *BioEssays* **25**, 174–181 (2003).
- Lu, K. P., Hanes, S. D. & Hunter, T. A human peptidyl-prolyl isomerase essential for regulation of mitosis. *Nature* **380**, 544–547 (1996).
- Yaffe, M. B. et al. Sequence-specific and phosphorylation-dependent proline isomerization: a potential mitotic regulatory mechanism. *Science* **278**, 1957–1960 (1997).
- Lu, K. P., Liou, Y. C. & Zhou, X. Z. Pinning down the proline-directed phosphorylation signaling. *Trends Cell Biol.* **12**, 164–172 (2002).
- Lu, P. J., Wulf, G., Zhou, X. Z., Davies, P. & Lu, K. P. The prolyl isomerase Pin1 restores the function of Alzheimer-associated phosphorylated tau protein. *Nature* **399**, 784–788 (1999).
- Zhou, X. Z. et al. Pin1-dependent prolyl isomerization regulates dephosphorylation of Cdc25C and tau proteins. *Mol. Cell* **6**, 873–883 (2000).
- Holzer, M. et al. Inverse association of Pin1 and tau accumulation in Alzheimer's disease hippocampus. *Acta Neuropathol.* **104**, 471–481 (2002).
- Davies, D. C., Horwood, N., Isaacs, S. L. & Mann, D. M. The effect of age and Alzheimer's disease on pyramidal neuron density in the individual fields of the hippocampal formation. *Acta Neuropathol.* **83**, 510–517 (1992).

22. Liou, Y. C. *et al.* Loss of Pin1 function in the mouse resembles the cyclin D1-null phenotypes. *Proc. Natl Acad. Sci. USA* **99**, 1335–1340 (2002).
23. Husseman, J. W., Nochlin, D. & Vincent, I. Mitotic activation: a convergent mechanism for a cohort of neurodegenerative diseases. *Neurobiol. Aging* **21**, 815–828 (2000).
24. Kins, S. *et al.* Reduced protein phosphatase 2A activity induces hyperphosphorylation and altered compartmentalization of tau in transgenic mice. *J. Biol. Chem.* **276**, 38193–38200 (2001).
25. Wolozin, B. L., Pruchnicki, A., Dickson, D. W. & Davies, P. A neuronal antigen in the brains of Alzheimer patients. *Science* **232**, 648–650 (1986).
26. Lee, V. M., Balin, B. J., Otvos, L. Jr & Trojanowski, J. Q. A68: a major subunit of paired helical filaments and derivatized forms of normal Tau. *Science* **251**, 675–678 (1991).
27. Gordon-Krajcer, W., Yang, L. & Ksiezak-Reding, H. Conformation of paired helical filaments blocks dephosphorylation of epitopes shared with fetal tau except Ser199/202 and Ser202/Thr205. *Brain Res.* **856**, 163–175 (2000).
28. Fujimori, F., Takahashi, K., Uchida, C. & Uchida, T. Mice lacking Pin1 develop normally, but are defective in entering cell cycle from G90 arrest. *Biochem. Biophys. Res. Commun.* **265**, 658–663 (1999).
29. Patrick, G. N. *et al.* Conversion of p35 to p25 deregulates Cdk5 activity and promotes neurodegeneration. *Nature* **402**, 615–622 (1999).
30. Lucas, J. J. *et al.* Decreased nuclear β -catenin, tau hyperphosphorylation and neurodegeneration in GSK-3 β conditional transgenic mice. *EMBO J.* **20**, 27–39 (2001).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank F. Gage, L. Cantley, B. Neel and K. Kosik for comments on the manuscript; W. Markesbery for human brain samples; P. Davies for tau antibodies; and G. Liu, M. Liu and M. Ericsson for technical assistance. Y.-C.L. is a Fellow of the Canadian Institutes of Health Research, A.R. is a Special Fellow of the Leukemia and Lymphoma Society; T.H. is a Frank and Elsie Schilling American Cancer Society Research Professor; and K.P.L. is a Pew Scholar and a Leukemia and Lymphoma Society Scholar. This study was supported by NIH grants to G.B., X.J.L., T.H. and K.P.L.

Competing interests statement The authors declare competing financial interests; details accompany the paper on www.nature.com/nature.

Correspondence and requests for materials should be addressed to K.P.L. (klu@bidmc.harvard.edu).

Essential role for the peroxiredoxin Prdx1 in erythrocyte antioxidant defence and tumour suppression

Carola A. Neumann¹, Daniela S. Krause¹, Christopher V. Carman¹, Shampa Das¹, Devendra P. Dubey¹, Jennifer L. Abraham¹, Roderick T. Bronson³, Yuko Fujiwara², Stuart H. Orkin² & Richard A. Van Etten^{1*}

¹Center for Blood Research and Department of Genetics, Harvard Medical School, and

²Howard Hughes Medical Institute, Children's Hospital, Boston, Massachusetts 02115, USA

³Tufts University School of Veterinary Medicine, North Grafton, Massachusetts 01536, USA

* Present address: Molecular Oncology Research Institute, Tufts-New England Medical Center, Boston, Massachusetts 02111, USA

Reactive oxygen species are involved in many cellular metabolic and signalling processes¹ and are thought to have a role in disease, particularly in carcinogenesis and ageing². We have generated mice with targeted inactivation of *Prdx1*, a member of the peroxiredoxin family of antioxidant enzymes³. Here we show that mice lacking *Prdx1* are viable and fertile but have a shortened lifespan owing to the development beginning at about 9 months of severe haemolytic anaemia and several malignant cancers, both of which are also observed at increased frequency in heterozygotes. The haemolytic anaemia is characterized by an increase in erythrocyte reactive oxygen species, leading to protein oxidation, haemoglobin instability, Heinz body formation and decreased erythrocyte lifespan. The malignancies include lymphomas, sarcomas and carcinomas, and are frequently associ-

ated with loss of *Prdx1* expression in heterozygotes, which suggests that this protein functions as a tumour suppressor. *Prdx1*-deficient fibroblasts show decreased proliferation and increased sensitivity to oxidative DNA damage, whereas *Prdx1*-null mice have abnormalities in numbers, phenotype and function of natural killer cells. Our results implicate *Prdx1* as an important defence against oxidants in ageing mice.

Cellular defences against reactive oxygen species (ROS) include enzymes such as superoxide dismutase (which converts superoxide to hydrogen peroxide), catalase and glutathione peroxidase (which convert hydrogen peroxide to water), as well as non-enzymatic scavengers such as glutathione, ascorbic acid and carotenoids. Peroxiredoxins (Prdxs), a family of small antioxidant proteins that contain essential catalytic cysteine residues and use thioredoxin as an electron donor³, also scavenge peroxide and are thought to be involved in the cellular response to ROS. Prdxs are abundant proteins found in organisms from all three kingdoms, with at least five distinct members in mammals. Mammalian *Prdx1*, also known as Pag⁴ or MSP23 (ref. 5), is a ubiquitously expressed protein with a relative molecular mass of 23,000 (23K) that is encoded by a single gene on human chromosome 1p34 (ref. 6) and mouse chromosome 4 (ref. 7) and induced by serum stimulation⁴ and oxidative stress^{5,8}. Transfection studies show that *Prdx1* can eliminate peroxide *in vivo* and can regulate ROS induced by growth factor signalling⁹. In addition to its role as an antioxidant enzyme, *Prdx1* has been independently isolated as an erythrocyte cytosolic protein that enhances the cytotoxicity of natural killer (NK) cells¹⁰, a

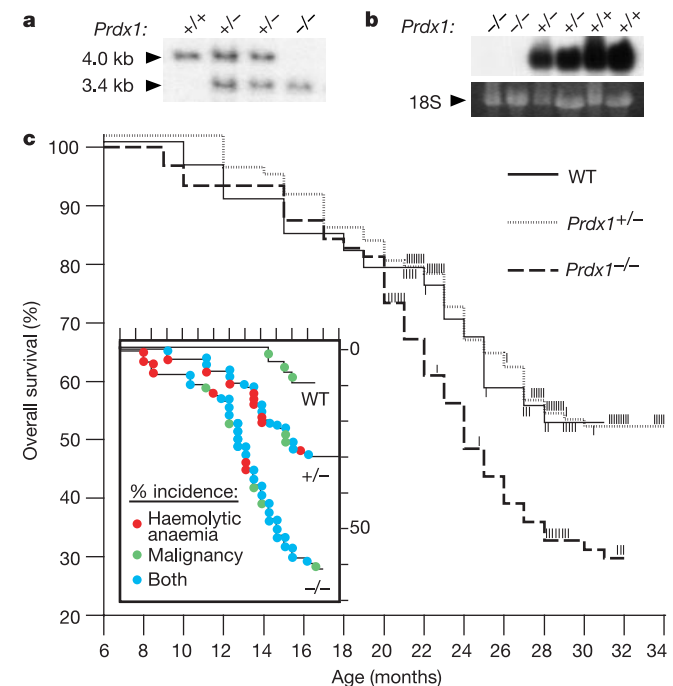


Figure 1 Premature death in ageing *Prdx1*^{-/-} mice. **a**, Genotype of four littermates from a *Prdx1*^{+/-} cross. Southern blot of *SpeI*-digested genomic DNA with a *Prdx1* exon II probe shows wild-type and mutant alleles of 4.0 and 3.4 kilobases, respectively. **b**, Northern blot of total liver RNA from six littermates hybridized with a *Prdx1* complementary DNA probe (top). Ethidium bromide staining of 18S rRNA (bottom) verifies equivalent loading. **c**, Kaplan-Meier survival curve of cohorts of wild-type ($n = 34$), *Prdx1*^{+/-} ($n = 88$) and *Prdx1*^{-/-} ($n = 64$) littermates on a mixed B6 $\times$ 129SvEv background. Mutant lines generated from three independently targeted ES cell clones were studied with similar results. The ages of surviving mice are indicated by tick marks. The difference in survival between wild-type and *Prdx1*^{-/-} mice is statistically significant ($P = 0.05$, Mantel-Cox test). Inset, percentage of mice in these cohorts that developed haemolytic anaemia (red), malignancy (green) or both (blue). The x axis is identical to the main graph.

Table 1 Blood parameters of wild-type and anaemic *Prdx1*^{-/-} mice

	Wild type	<i>Prdx1</i> ^{-/-}
Erythrocytes ($\times 10^6$ per μ l)	8.7 $\pm$ 0.2	2.4 $\pm$ 0.7*
Haematocrit (%)	42.1 $\pm$ 2.1	13.1 $\pm$ 0.1*
Haemoglobin (g dl ⁻¹)	13.4 $\pm$ 0.5	4.1 $\pm$ 0.6*
Erythrocyte volume distribution width (%)	14.2 $\pm$ 0.2	24.5 $\pm$ 2.0*
Reticulocytes (%)	3.5 $\pm$ 2.0	34.0 $\pm$ 9.9*
White blood cells ($\times 10^3$ per μ l)	8.5 $\pm$ 2.7	10.2 $\pm$ 9.5
Platelets ($\times 10^3$ per μ l)	1,185 $\pm$ 169	773 $\pm$ 94
Lactate dehydrogenase (U dl ⁻¹)	270 $\pm$ 21	842 $\pm$ 229*
Haptoglobin (U dl ⁻¹)	12 $\pm$ 1	<7*
Spleen weight (g)	0.2 $\pm$ 0.1	1.0 $\pm$ 0.6*

Peripheral blood was obtained from anaemic (defined as haemoglobin < 9) *Prdx1*^{-/-} mice ($n = 8$) and age-matched wild-type littermate controls ($n = 6$). Spleen weights were recorded at autopsy. Values are the mean $\pm$ s.e.m.

*For all parameters tested except white blood cells and platelets, the difference between wild-type and *Prdx1*^{-/-} mice was significant ($P < 0.01$, unpaired t -test).

cytosolic protein from hepatocytes with high affinity for haem¹¹, and a cytosolic and nuclear protein that functions as a stoichiometric inhibitor of the non-receptor tyrosine kinase c-Abl¹².

To determine the biological roles of *Prdx1*, we inactivated the *Prdx1* gene by homologous recombination in murine embryonic stem (ES) cells (Supplementary Fig. 1) and generated heterozygous *Prdx1*^{+/-} mice. In crosses of *Prdx1*^{+/-} mice, *Prdx1*^{-/-} mice were born at the expected mendelian frequency (Fig. 1a and data not shown), lacked *Prdx1* messenger RNA (Fig. 1b), and showed normal postnatal development and fertility. However, prolonged observation of cohorts of wild-type, *Prdx1*^{-/-} and *Prdx1*^{+/-} littermates showed a significantly shortened survival of *Prdx1*^{-/-} mice relative to wild-type littermates (Fig. 1c). Clinicopathological analysis suggested that there were two main causes of premature death in ageing *Prdx1*^{-/-} mice: haemolytic anaemia, which first appeared at 9 months; and malignant tumours. Both diseases were also observed at increased frequency in *Prdx1*^{+/-} mice, beginning at around 12 months, whereas no wild-type mice developed haemolytic anaemia (Fig. 1c, inset).

Many premorbid *Prdx1*^{-/-} mice had severe anaemia characterized by a marked decrease in haematocrit and haemoglobin in peripheral blood relative to wild-type littermate controls, with normal leukocyte and platelet counts (Table 1). Peripheral blood smears from these mice showed prominent microcytosis, anisocytosis, poikilocytosis and polychromatophilia of erythrocytes (Fig. 2a, b), which coincided with an increase in both the width of the erythrocyte volume distribution (Table 1) and the numbers of reticulocytes (Fig. 2c and Table 1). At autopsy, anaemic mice frequently had splenomegaly (Table 1), sometimes massive, owing to extramedullary erythropoiesis (Fig. 2d). These findings are suggestive of anaemia due to decreased survival, rather than impaired production, of erythrocytes. In agreement with this, anaemic mice had increased blood lactate dehydrogenase and decreased haptoglobin (Table 1), indicative of the intravascular destruction of red cells. Similar clinicopathological features were observed in *Prdx1*^{+/-} mice (data not shown), which also developed severe anaemia with age (Fig. 1d).

Serial analysis of blood haemoglobin in the *Prdx1*^{-/-} mice showed a general and gradual decline, which began at around 12 months without any change in reticulocyte numbers (Fig. 2e), similar to wild-type mice (data not shown). However, many ageing *Prdx1*^{-/-} mice showed precipitous drops in blood haemoglobin over an interval of 1–2 months, coincident with and in some mice preceded by an increase in circulating reticulocytes. These data suggest that a subset of *Prdx1*^{-/-} mice develop haemolytic anaemia as they age.

To investigate the cause of the anaemia, we compared the survival of biotin-labelled erythrocytes obtained from anaemic *Prdx1*^{-/-} mice against those from healthy wild-type littermates on transfer into wild-type recipients (Fig. 2f). Whereas the rate of disappearance of labelled wild-type erythrocytes was roughly linear, with a loss of about 2% per day, *Prdx1*^{-/-} erythrocytes disappeared at an exponential rate, indicative of an increase in elimination *in vivo* that was independent of red cell age. These results show that the anaemia

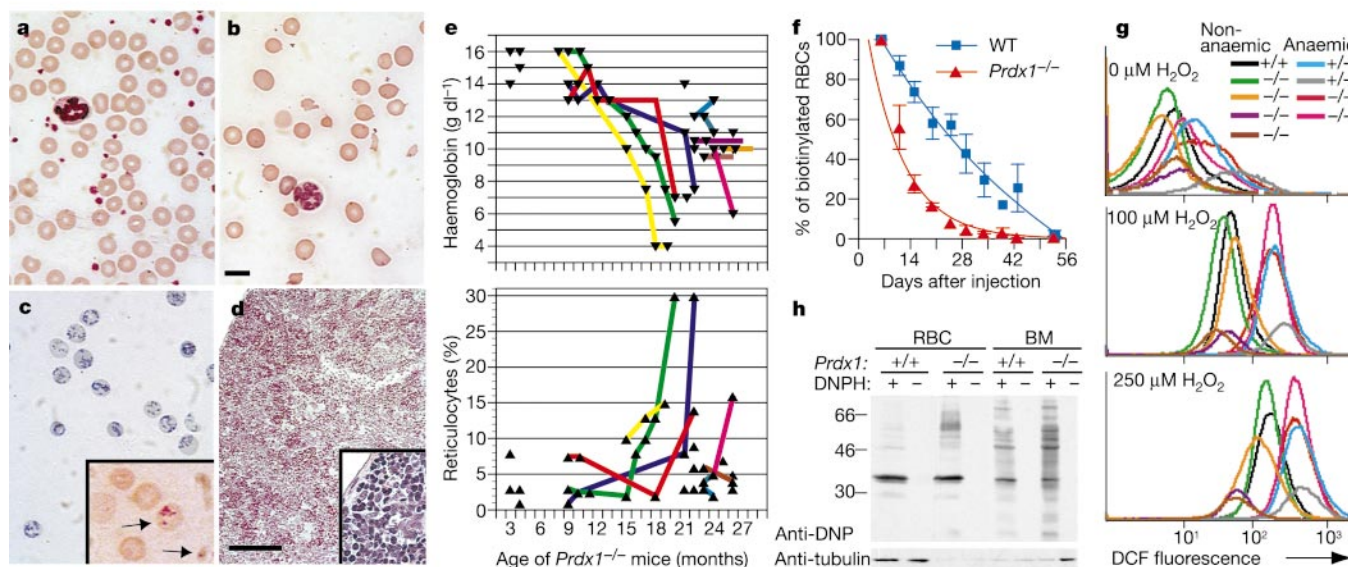


Figure 2 Haemolytic anaemia caused by intra-erythrocytic oxidative damage in *Prdx1* mutant mice. **a, b**, Peripheral blood smear (Wright–Giemsa stain) from a healthy wild-type littermate (**a**) and a *Prdx1*^{-/-} mouse with severe anaemia (**b**). **c**, Peripheral blood (methylene blue stain) from an anaemic *Prdx1*^{-/-} mouse with increased numbers of reticulocytes. Inset, methyl violet stain showing oxidized haemoglobin precipitates (Heinz bodies, arrows). **d**, Spleen (haematoxylin and eosin stain) from an anaemic *Prdx1*^{-/-} mouse, showing disruption of the follicular architecture by extensive extramedullary erythropoiesis (inset). **e**, Blood haemoglobin (top) and reticulocyte counts (bottom) in a subset of *Prdx1*^{-/-} mice plotted against age. Coloured lines link serial values from a single mouse. **f**, Decrease in the intrinsic survival of biotin-labelled erythrocytes from anaemic *Prdx1*^{-/-} mice after adoptive transfer to wild-type recipients. The half-lives

were 25 d for wild-type (squares) and 10 d for *Prdx1*^{-/-} (triangles) erythrocytes. Bars indicate the s.e.m. from quadruplicate recipients of the same donor sample; results are representative of two independent experiments. **g**, ROS concentrations measured by DCF fluorescence in erythrocytes from healthy (non-anaemic) wild-type and *Prdx1*^{-/-} mice and from anaemic *Prdx1*^{+/-} and *Prdx1*^{-/-} mice exposed to peroxide *in vitro*. **h**, Protein oxidation in erythrocytes (RBC) and bone marrow cells (BM) from wild-type and anaemic *Prdx1*^{-/-} mice detected by western blotting with an antibody against DNP (top) after treatment with DNPH. After correcting for protein levels (bottom), there was a 6.9-fold and 1.5-fold increase in oxidized proteins in *Prdx1*^{-/-} RBCs and BMs, respectively. Scale bars, 10 μ m (**a–c**); 100 μ m (**d**).

in ageing *Prdx1*^{-/-} mice is due to shortened erythrocyte survival owing to an intrinsic defect in red cells.

The presence of Heinz bodies, representing precipitated haemoglobin, in *Prdx1*^{-/-} erythrocytes (Fig. 2c, inset) suggested that the cause of red cell destruction might be increased erythrocyte ROS, leading to the oxidation of red cell proteins including haemoglobin. In support of this, we observed an increase both in baseline ROS and in ROS generated in response to hydrogen peroxide challenge in erythrocytes from anaemic *Prdx1*^{-/-} and *Prdx1*^{+/-} mice but not in those from age-matched healthy *Prdx1* mutant or wild-type mice (Fig. 2g). Analysis of the carbonyl groups in cellular proteins, a by-product of oxidation, confirmed the increased oxidation of polypeptides in erythrocytes and bone marrow cells of anaemic *Prdx1* mutant mice (Fig. 2h). Finally, we confirmed the presence of oxidized, unstable haemoglobin in erythrocytes from anaemic *Prdx1*^{-/-} mice by showing the increased sensitivity of free haemoglobin in red cell lysates to precipitation by isopropanol (mean precipitated haemoglobin, 33.5 ± 4.3% for *Prdx1*^{-/-} erythrocytes versus 2.1 ± 2.0% for wild type; *P* < 0.01, unpaired *t*-test). Collectively, these results show that ageing *Prdx1* mutant mice develop fatal haemolytic anaemia owing to an increase in erythrocyte ROS, oxidation and precipitation of haemoglobin, and haemolysis.

Cancer was a second main cause of morbidity and death in ageing *Prdx1* mutant mice. Only 3 of 34 wild-type mice died with evidence of malignancy: two from disseminated histiocytic tumours and one from B-cell lymphoma. Both of these cancers are typically found in ageing B6, 129SvEv mice. By contrast, half (32/64) of the *Prdx1*^{-/-} mice succumbed to malignancy either alone or in combination with haemolytic anaemia (Fig. 1d). In addition to B and T lymphomas and histiocytic malignancy, the spectrum of cancers that developed in *Prdx1* mutant mice included epithelial and mesenchymal tumours (hepatocellular carcinoma, fibrosarcoma, osteosarcoma,

islet cell adenomas and adenocarcinomas of lung and breast), which are less common in ageing B6, 129SvEv mice (Fig. 3a and Supplementary Fig. 2).

Most sarcomas and lymphomas could be transplanted into nude or sublethally irradiated (275 cGy) severe combined immunodeficient (SCID) mice with latencies of around 1 month or 6 months, respectively (data not shown). An increase in the incidence of malignancy was also observed in *Prdx1*^{+/-} mice (Figs 1d and 3a), although this was not associated with a decrease in overall survival. It is likely that some mice in this cohort had malignancy as a co-morbid condition but died of other causes such as infection. Although the expression of *Prdx1* protein was abundant in many normal mouse tissues (Fig. 3b), it was low to undetectable in lysates from several different tumours isolated from *Prdx1*^{+/-} mice (Fig. 3c). This is reminiscent of the inactivation of a tumour suppressor gene and suggests that loss of *Prdx1* function may contribute to tumorigenesis. We did not observe any gross structural rearrangements or loss of the wild-type *Prdx1* allele in any tumours from *Prdx1*^{+/-} mice (data not shown), suggesting that either subtle mutations or epigenetic mechanisms such as methylation might be responsible.

To investigate possible mechanisms of tumorigenesis, we characterized murine embryonic fibroblasts (MEFs) derived from *Prdx1*^{-/-} and wild-type littermates. *Prdx1*^{-/-} MEFs proliferated more slowly and had a higher fraction of cells in the G1 phase of the cell cycle than did wild-type MEFs (Fig. 3d). A role for *Prdx1* in the cell cycle has been previously suggested because *Prdx1* is phosphorylated by Cdc2 during mitosis, resulting in an inhibition of *Prdx1* peroxidase activity¹³. The mechanism underlying the slower proliferation of cells lacking *Prdx1* is unknown but might reflect chronic oxidative stress. Other studies have shown that overexpressing peroxiredoxins increases cellular resistance to oxidative stress^{14,15}. Consistent with this, *Prdx1*^{-/-} MEFs showed signifi-

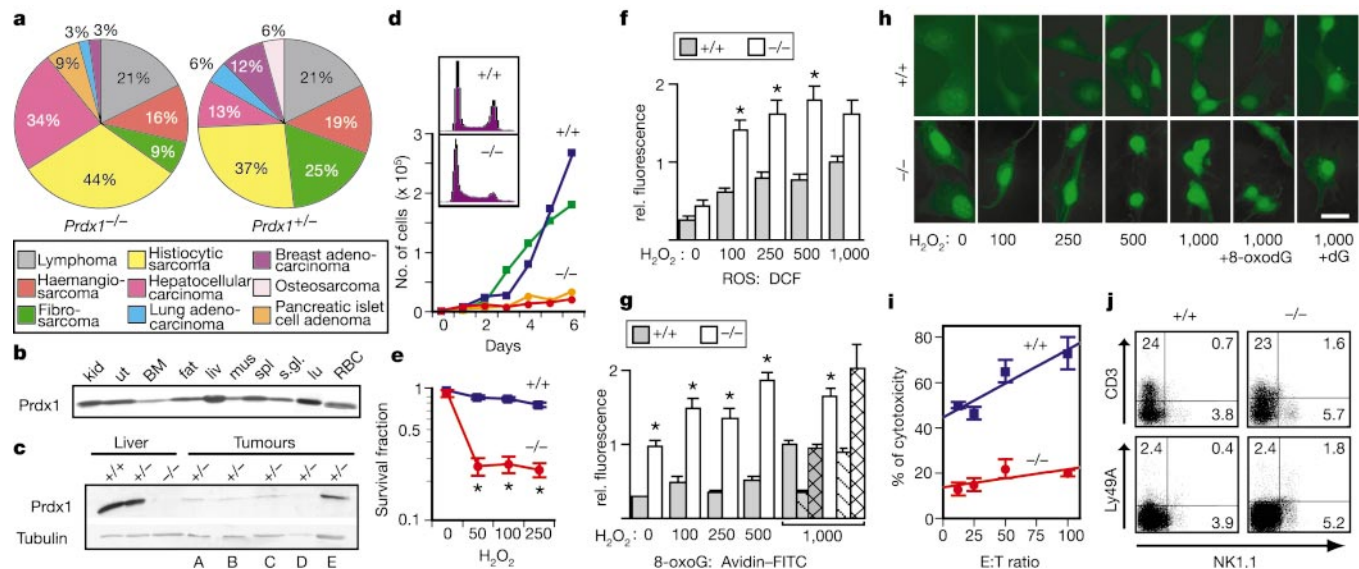


Figure 3 *Prdx1* mutant mice are predisposed to cancer. **a**, Histopathological distribution of tumours observed in *Prdx1*^{-/-} (*n* = 32) and *Prdx1*^{+/-} (*n* = 16) mice. Whereas 38% of tumour-bearing *Prdx1*^{-/-} and *Prdx1*^{+/-} mice had two independent malignancies, 6% of *Prdx1*^{-/-} mice were diagnosed with three different tumours. **b**, Tissue lysates analysed by western blotting with antisera against a GST-*Prdx1* fusion protein. **c**, Lysates from normal liver of wild-type, *Prdx1*^{+/-} or *Prdx1*^{-/-} mice (left three lanes) and from independent tumours of *Prdx1*^{+/-} mice (right five lanes: A and B, haemangiosarcoma; C and E, fibrosarcomas; D, osteosarcoma) analysed by western blotting with antibodies against *Prdx1* (top) or tubulin (bottom). **d**, Proliferation of *Prdx1*^{-/-} (squares) and wild-type (circles) MEFs. Inset, the fraction of cells with a 2*n* DNA content averaged 44% in wild-type MEFs (top) and 61% in *Prdx1*^{-/-} MEFs (bottom). **e**, Clonogenic survival of *Prdx1*^{-/-} (squares) and wild-type (circles) MEFs after oxidative stress (mean ± s.e.m.

fractional survival). **f**, Basal and peroxide-induced ROS in *Prdx1*^{-/-} (white bars) and wild-type (grey bars) MEFs, measured by DCF and expressed as the normalized mean ± s.e.m. intensity of cell fluorescence. **g**, Basal and peroxide-induced 8-oxoG in *Prdx1*^{-/-} and wild-type MEFs, assessed by staining with avidin-FITC and expressed as in **f**. To show specificity, avidin-FITC was preincubated with an oligonucleotide containing a single 8-oxodeoxyguanosine residue (hatched bars) or a control unoxidized oligonucleotide (crosshatched bars). In **e–g**, **P* < 0.001 (unpaired *t*-test). **h**, Representative photomicrographs of the wild-type and *Prdx1*^{-/-} MEFs shown in **g** treated with the indicated concentration of peroxide and stained with avidin-FITC. Scale bar, 20 μm. **i**, Cytotoxic activity of splenic NK cells from *Prdx1*^{-/-} (circles) and wild-type (squares) mice; bars indicate the s.d. **j**, Frequency of CD3⁺NK1.1⁺ and NK1.1⁺Ly49A⁺ cells in enriched splenic NK cell populations from *Prdx1*^{-/-} and wild-type mice.

cantly lower clonogenic survival in response to oxidant treatment (Fig. 3e) and had greater concentrations of peroxide-induced cellular ROS (Fig. 3f) as compared with wild-type MEFs.

We also examined cellular oxidative DNA damage by fluorescence staining of MEFs with fluorescein isothiocyanate (FITC)-conjugated avidin, which detects 8-oxoguanine (8-oxoG)¹⁶—a principal oxidative DNA lesion that can cause base mispairing and mutations. *Prdx1*^{−/−} MEFs had significantly greater basal and peroxide-induced concentrations of 8-oxoguanine (Fig. 3g) than did wild-type MEFs. The staining was predominantly nuclear (Fig. 3h) and was significantly blocked by preincubating avidin-FITC with an 8-oxoG-containing oligonucleotide but not a non-oxidized oligonucleotide (Fig. 3g, h). Collectively, these observations suggest that loss of Prdx1 increases the susceptibility of non-erythroid tissues to cancer by causing a heightened sensitivity to oxidants and an increase in both cellular ROS and oxidative DNA damage. The loss of Prdx1 protein expression in tumours from *Prdx1*^{+/−} mice is consistent with this mechanism.

Natural killer cells are lymphocyte effector cells of the innate immune system that may be important in protecting against tumour development¹⁷, and Prdx1 has been identified as an erythrocyte cytosolic protein, NK enhancing factor A (NKEF-A), that stimulates NK activity¹⁰. We therefore examined NK cells of *Prdx1* mutant mice as another mechanism that might contribute to tumorigenesis. Partially purified splenic NK cells from *Prdx1*^{−/−} mice showed a reproducible and statistically significant decrease in lytic activity towards YAC-1 target cells (Fig. 3i). Flow cytometric analysis detected no significant difference in the frequency of splenic CD3⁺NK1.1[−] (19.6 ± 1.9% versus 15.9 ± 3%, respectively) and CD3⁺ NK1.1⁺ cells (0.6 ± 0.2% versus 0.8 ± 0.2%, respectively) between wild-type and *Prdx1*^{−/−} mice; however, an increase (40% to 90%) in the frequency of CD3[−]NK1.1⁺ cells was observed in all *Prdx1*^{−/−} mice tested (Fig. 3j). A two- to threefold increase in the frequency of a subset of NK cells expressing the inhibitory receptor Ly49A (NK1.1⁺Ly49A⁺ cells)¹⁸ was also observed in *Prdx1*^{−/−} mice (Fig. 3j). The frequency of NK cells expressing the activation receptor Ly49D¹⁸ was more variable, but in three out of four experiments we observed reduced (<80% of wild-type) frequencies of NK1.1⁺Ly49D⁺ cells in *Prdx1*^{−/−} mice (data not shown).

We confirmed previous observations that addition of erythrocytes stimulates the cytotoxic activity of NK cells¹⁰, and observed a significant decrease in red blood cell (RBC) NK-enhancing activity when erythrocytes from *Prdx1*^{−/−} mice were added to NK cells from wild-type mice (Table 2). The residual NK-enhancing activity of RBCs lacking Prdx1 could be due to the presence of Prdx2 (also known as NKEF-B), although only Prdx1 shows NK-enhancing activity as a purified protein¹⁹. Notably, wild-type RBCs had little stimulatory effect on the minimal cytotoxic activity of *Prdx1*^{−/−} NK cells (Table 2), suggesting that Prdx1 may be required in both NK cells and non-immune cells for optimal NK function *in vivo*.

Our results implicate Prdx1 directly in protecting against ROS in red cells and in tumour suppression in ageing mice. Physiological roles for peroxiredoxins in both areas have been postulated²⁰, but

direct proof has been lacking. The main antioxidant defences in red cells were thought to be the glutathione peroxidase and catalase enzyme systems, which use nicotinamide adenine dinucleotide phosphate (NADPH) generated by the hexose monophosphate shunt through glucose 6-phosphate dehydrogenase as the electron donor and glutathione as the direct ROS scavenger. But an absence of either catalase or glutathione peroxidase-1 (ref. 21) does not lead to haemolysis in mice, whereas deficiency of haematopoietic mitochondrial superoxide dismutase causes only mild haemolysis that does not impair survival²². By contrast, loss of Prdx1 causes severe fatal haemolytic anaemia but only in ageing mice, which suggests that other red cell antioxidant defences initially compensate for the lack of Prdx1 but that deficiency in these pathways acquired with age leads to haemolysis. Notably, *Prdx1*^{+/−} mice with haemolysis still express Prdx1 in bone marrow or red cells but have about 50% of the Prdx1 concentrations of wild-type mice in these tissues (data not shown), suggesting that large amounts of Prdx1 are necessary for normal red cell survival in older mice.

Similarly, although a role for ROS in carcinogenesis has long been postulated, several mutant mouse strains with deficiencies in antioxidant pathways show an increase in spontaneous cancer development only in response to carcinogens^{23,24}. Our results argue that Prdx1 has a direct role in tumour suppression in older mice by eliminating ROS and preventing oxidative DNA damage. Prdx1 and other peroxiredoxins are overexpressed in some human cancers²⁵, suggesting that tumours that arise through other mechanisms may benefit from increased amounts of peroxiredoxins. Our findings further suggest that Prdx1 regulates NK cell development and cytotoxic function through both cell-autonomous and cell-non-autonomous mechanisms, and that defective NK cell activity may predispose *Prdx1* mutant mice to cancer through a loss of tumour surveillance.

Prdx1 can also inhibit the function of both c-Abl¹² and c-Myc²⁶, two proteins whose constitutively active forms cause several haematopoietic neoplasms including lymphoma and histiocytic sarcoma. We did not investigate Myc levels or function in *Prdx1*^{−/−} tumours, but several tumours showed an increase in protein tyrosine phosphorylation relative to normal tissues. We did not detect direct tyrosine phosphorylation of c-Abl in these tumours by immunoprecipitation and western blot analysis (data not shown), but this may in part reflect redundancy among the peroxiredoxins, because Prdx3 can also inhibit Abl in transfection experiments (C.A.N. and R.A.V., unpublished data). Establishing the precise mechanisms underlying the cancer susceptibility of *Prdx1* mutant mice will require further studies. These mice should be a valuable tool for understanding the role of antioxidant pathways in ageing, carcinogenesis and other pathophysiological processes.

Note added in proof: Mice deficient in Peroxiredoxin 2 were recently reported to develop nonfatal haemolytic anemia³¹. □

Methods

Generation of *Prdx1*^{−/−} mice

An *Avr1* fragment from a bacterial artificial clone spanning the *Prdx1* locus was cloned into the vector pCWO. A targeting vector was generated by transposon-mediated mutagenesis²⁷, in which a transposon carrying unique restriction sites was inserted into exon III of *Prdx1* (Supplementary Fig. 1). We then cloned a *PGK-neo* expression cassette, consisting of the neomycin resistance gene fused to *PGK* promoter, into this vector. ES cells (line Tc-1) were electroporated with this construct and selected for resistance to G418 and ganciclovir. Southern blot analysis showed that 52 of 129 doubly resistant clones analysed were correctly targeted. We injected three independent clones into blastocysts from B6 mice and crossed the resulting chimeras back to B6 mice to generate *Prdx1*^{+/−} mice.

Erythrocyte survival, ROS, protein oxidation and unstable haemoglobin

Erythrocytes from individual healthy wild-type and anaemic *Prdx1*^{−/−} mice were labelled *in vivo* with biotin-X-N-hydroxysuccinimide ester (Calbiochem) as described²⁸, 10⁷ labelled erythrocytes were injected intravenously into wild-type recipient mice, and the kinetics of the disappearance of biotin-labelled cells from circulation was measured by flow cytometric staining with phycoerythrin-conjugated streptavidin.

To measure ROS, we loaded erythrocytes with 10 μM 2', 7'-dichlorodihydrofluorescein diacetate (DCF, Sigma) and analysed intracellular fluorescence intensity by flow cytometry

Table 2 Effects of RBC addition and *Prdx1* genotype on NK cytotoxic activity

Fold increase in cytotoxic activity*					Effector to RBC ratio
<i>Prdx1</i> genotype of cells					
NK cells:	+/+	+/+	-/-	-/-	
RBCs:	+/+	-/-	+/+	-/-	
	5.3 ± 1.5†	3.0 ± 1.8	1.2 ± 0.5	1.0 ± 0.5	1:5
	25.6 ± 2.7	16.4 ± 1.7	2.6 ± 0.4	2.9 ± 0.4	1:20

Washed RBCs isolated from wild-type and *Prdx1*^{−/−} mice were mixed with effector cells at two concentrations (RBC to effector ratios of 5 and 20) and a 4-h chromium release assay was done as described³⁰.

*The effect of RBCs on NK activity is expressed as the fold increase in lytic activity (LA), calculated as (LA in the presence of RBCs − LA in the absence of RBCs)/(LA in the absence of RBCs). RBCs alone did not induce target cell lysis. Values are the mean ± s.e.m.

†Relative to *Prdx1*^{−/−} cells, wild-type NK cells and RBCs showed a 5- and 9-fold increase in lytic activity at RBC to effector ratios of 5 and 20, respectively.

as described²⁹. Protein oxidation in erythroid tissues from anaemic *Prdx1*^{-/-} and age-matched wild-type mice was determined by incubating lysates in the presence or absence of 2,4-dinitrophenylhydrazine (DNPH) to derivatize oxidized carbonyl groups to 2,4-dinitrophenylhydrazone²⁹, followed by western blotting with an antibody against DNP (Invitrogen). To measure unstable haemoglobin, we lysed erythrocytes from anaemic *Prdx1*^{-/-} (*n* = 3) and age-matched wild-type control (*n* = 4) mice with distilled water and precipitated the membranes by adding 150 mM NaCl. The supernatant was analysed for haemoglobin concentration before and after the addition of 17% isopropanol by spectrophotometric determination of absorbance at 540–580 nm.

Cell cycle, clonogenic survival, ROS and 8-oxoG analysis

MEFs established from two independent pairs of wild-type and *Prdx1*^{-/-} littermates were plated in duplicate cultures and the cell number was determined daily. We analysed the DNA content by flow cytometric analysis using propidium iodide staining of unsynchronized cells collected on day 4. To assess clonogenic survival in response to oxidative stress, MEFs were treated with hydrogen peroxide for 30 min and plated in triplicate, and the colonies were counted 7 d later.

To measure ROS concentrations, MEFs were stimulated with peroxide for 10 min in serum- and phenol-red-free medium, incubated with 50 μ M DCF for 10 min, and then washed and analysed with a Axiovert S200 microscope (Zeiss) and an Ocrs CCD (charge-coupled device) camera (Hamamatsu). At least 15 individual cell images were acquired per condition, background levels were subtracted, and the mean total cellular fluorescence intensity was quantified by OpenLab 3.1.2 software (Improvision). Results are expressed as the mean fluorescence intensity normalized to that of wild-type cells treated with 1,000 μ M peroxide.

To measure 8-oxoG, MEFs growing on fibronectin-coated coverslips were treated with peroxide in serum-free and phenol-red-free medium for 2 h at 37 °C, fixed in absolute methanol (-20 °C, 20 min) and permeabilized with 0.1% Triton X-100 (room temperature, 15 min). After nonspecific binding sites were blocked, the cells were stained with 15 μ g ml⁻¹ FITC-conjugated avidin (Sigma) for 1 h at 37 °C. To verify specificity of the detection of 8-oxoG, avidin-FITC was preincubated with an eightfold excess of either a 23-base oligodeoxynucleotide (Sigma) containing a single 8-oxodeoxyguanosine residue or a control unoxidized oligonucleotide. We measured and quantified the normalized the mean cellular fluorescence intensity as described above.

NK cell cytolytic activity and cell-surface antigen expression

Splenocytes from *Prdx1*^{-/-} (*n* = 15) and age-matched wild-type (*n* = 15) young mice were enriched for NK cells by centrifugation over Lympholyte-M (Cedarlane Laboratories) and passage over a nylon wool column. We determined cytotoxic activity in duplicate at four different effector to target ratios in a 4-h chromium-release assay by using ⁵¹Cr-labelled YAC-1 cells as a target as described³⁰. Specific cytotoxicity was calculated as the percentage of specific cytotoxicity = [c.p.m.(ER) - c.p.m.(SR)]/[c.p.m.(MR) - c.p.m.(SR)] $\times$ 100, where ER is the ⁵¹Cr release in experimental wells, SR is the spontaneous ⁵¹Cr release, and MR is the maximum ⁵¹Cr release in the presence of 10% SDS. Cytotoxic activity was expressed in lytic activity units (LA) per 10⁶ cells, which is defined as the number of effector cells required for 20% lysis of the labelled target cells. The specific cytotoxic activity of wild-type and *Prdx1*^{-/-} NK cells was 91 and 3.6 LA per 10⁶ cells, respectively. Enriched splenic NK cell populations from wild-type and *Prdx1*^{-/-} mice were analysed by flow cytometry with directly conjugated monoclonal antibodies against CD3, NK1.1, Ly49A and Ly49D (PharMingen).

Received 14 February; accepted 27 May 2003; doi:10.1038/nature01819.

- Finkel, T. Oxygen radicals and signaling. *Curr. Opin. Cell Biol.* **10**, 248–253 (1998).
- Finkel, T. & Holbrook, N. J. Oxidants, oxidative stress and the biology of ageing. *Nature* **408**, 239–246 (2000).
- Chae, H. Z. *et al.* Cloning and sequencing of thiol-specific antioxidant from mammalian brain: alkyl hydroperoxide reductase and thiol-specific antioxidant define a large family of antioxidant enzymes. *Proc. Natl Acad. Sci.* **91**, 7017–7021 (1994).
- Prosperi, M. T., Ferbus, D., Karczinski, I. & Goubin, G. A human cDNA corresponding to a gene overexpressed during cell proliferation encodes a product sharing homology with amoebic and bacterial proteins. *J. Biol. Chem.* **268**, 11050–11056 (1993).
- Ishii, T. *et al.* Cloning and characterization of a 23-kDa stress-induced mouse peritoneal macrophage protein. *J. Biol. Chem.* **268**, 18633–18636 (1993).
- Prosperi, M. T., Apiou, E., Dutrillaux, B. & Goubin, G. Organization and chromosomal assignment of two human PAG gene loci: PAGA encoding a functional gene and PAGB a processed pseudogene. *Genomics* **19**, 236–241 (1994).
- Lyu, M. S. *et al.* Genetic mapping of six mouse peroxiredoxin genes and fourteen peroxiredoxin related sequences. *Mammal. Genome* **10**, 1017–1019 (1999).
- Prosperi, M. T., Ferbus, D., Rouillard, D. & Goubin, G. The pag gene product, a physiological inhibitor of c-abl tyrosine kinase, is overexpressed in cells entering S phase and by contact with agents inducing oxidative stress. *FEBS Lett.* **423**, 39–44 (1998).
- Kang, S. W. *et al.* Mammalian peroxiredoxin isoforms can reduce hydrogen peroxide in response to growth factors and tumor necrosis factor- α . *J. Biol. Chem.* **273**, 6297–6302 (1998).
- Shau, H., Gupta, R. K. & Golub, S. H. Identification of a natural killer enhancing factor (NKEF) from human erythroid cells. *Cell. Immunol.* **147**, 1–11 (1993).
- Iwahara, S. *et al.* Purification, characterization, and cloning of a heme-binding protein (23 kDa) in rat liver cytosol. *Biochemistry* **34**, 13398–13406 (1995).
- Wen, S.-T. & Van Etten, R. A. The PAG gene product, a stress-induced protein with antioxidant properties, is an Abl SH3-binding protein and a physiological inhibitor of c-Abl tyrosine kinase activity. *Genes Dev.* **11**, 2456–2467 (1997).
- Chang, T.-S. *et al.* Regulation of peroxiredoxin I activity by Cdc2-mediated phosphorylation. *J. Biol. Chem.* **277**, 25370–25376 (2002).
- Shau, H. *et al.* Endogenous natural killer enhancing factor-B increases cellular resistance to oxidative

stress. *Free Radic. Biol. Med.* **22**, 497–507 (1997).

- Chung, Y. M., Yoo, Y. D., Park, J. K., Kiim, Y.-T. & Kim, H. J. Increased expression of peroxiredoxin II confers resistance to cisplatin. *Anticancer Res.* **21**, 1129–1134 (2001).
- Struthers, L., Patel, R., Clark, J. & Thomas, S. Direct detection of 8-oxodeoxyguanosine and 8-oxoguanine by avidin and its analogues. *Anal. Biochem.* **255**, 20–31 (1998).
- Cerwenka, A. & Lanier, L. L. Natural killer cells, viruses and cancer. *Nature Rev. Immunol.* **1**, 41–49 (2001).
- Takei, F., Brennan, J. & Mager, D. L. The Ly49 family: genes, proteins and recognition of class I MHC. *Immunol. Rev.* **155**, 67–77 (1997).
- Sauri, H., Ashjian, P. H., Kim, A. T. & Shau, H. Recombinant natural killer enhancing factor augments natural killer cytotoxicity. *J. Leuk. Biol.* **59**, 925–931 (1996).
- Butterfield, L. H., Merino, A., Golub, S. H. & Shau, H. From cytoprotection to tumor suppression: the multifactorial role of peroxiredoxins. *Antioxid. Redox Signal.* **1**, 385–402 (1999).
- Johnson, R. M., Goyette, G., Ravindranath, Y. & Ho, Y.-S. Red cells from glutathione peroxidase-1-deficient mice have nearly normal defenses against exogenous peroxides. *Blood* **96**, 1985–1988 (2000).
- Friedman, J. S. *et al.* Absence of mitochondrial superoxide dismutase results in a murine hemolytic anemia responsive to therapy with a catalytic antioxidant. *J. Exp. Med.* **193**, 925–934 (2001).
- Henderson, C. J. *et al.* Increased skin tumorigenesis in mice lacking pi class glutathione S-transferases. *Proc. Natl Acad. Sci. USA* **95**, 5275–5280 (1998).
- Noh, D.-Y. *et al.* Overexpression of peroxiredoxin in human breast cancer. *Anticancer Res.* **21**, 2085–2090 (2001).
- Mu, Z. M., Yin, X. Y. & Prochownik, E. V. Pag, a putative tumor suppressor, interacts with the Myc box II domain of c-Myc and selectively alters its biological function and target gene expression. *J. Biol. Chem.* **277**, 43175–43184 (2002).
- Westphal, C. H. & Leder, P. Transposon-generated 'knock-out' and 'knock-in' gene-targeting constructs for use in mice. *Curr. Biol.* **7**, 530–533 (1997).
- Hoffmann-Fezer, G. *et al.* Biotin labeling as an alternative nonradioactive approach to determination of red cell survival. *Ann. Hematol.* **67**, 81–87 (1993).
- Frank, J., Biesalski, H. K., Dominici, S. & Pompella, A. The visualization of oxidant stress in tissues and isolated cells. *Histol. Histopathol.* **15**, 173–184 (2000).
- Dubey, D. P. *et al.* The MHC influences NK and NKT cell functions associated with immune abnormalities and life span. *Mech. Aging Dev.* **113**, 117–134 (2000).
- Lee, T.-H. *et al.* Peroxiredoxin II is essential for sustaining life span of erythrocytes in mice. *Blood* **101**, 5033–5038 (2003).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank C. Westphal, P. Leder, K. Smith, J. Alvarez, J. Pinkas and C. Brugnara for assistance, and H. F. Bunn for discussions. This work was supported by NIH grants to C.A.N., S.D., D.P.D. and R.A.V.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.A.V. (rvanetten@tufts-nemc.org).

Machinery for protein sorting and assembly in the mitochondrial outer membrane

Nils Wiedemann^{1,2}, Vera Kozjak^{1,2}, Agnieszka Chacinska¹, Birgit Schönfisch¹, Sabine Rospert³, Michael T. Ryan^{1*}, Nikolaus Pfanner¹ & Chris Meisinger¹

¹Institut für Biochemie und Molekularbiologie, Universität Freiburg, Hermann-Herder-Str. 7, D-79104 Freiburg, Germany

²Fakultät für Biologie, Universität Freiburg, D-79104 Freiburg, Germany

³Max-Planck Research Unit Enzymology of Protein Folding, Weinbergweg 22, D-06120 Halle, Saale, Germany

* Present address: Department of Biochemistry, La Trobe University, 3086 Melbourne, Australia

Mitochondria contain translocases for the transport of precursor proteins across their outer and inner membranes^{1–5}. It has been assumed that the translocases also mediate the sorting of proteins to their submitochondrial destination^{1,2,5–10}. Here we show that the mitochondrial outer membrane contains a separate sorting and assembly machinery (SAM) that operates after the

translocase of the outer membrane (TOM). Mas37 forms a constituent of the SAM complex. The central role of the SAM complex in the sorting and assembly pathway of outer membrane proteins explains the various pleiotropic functions that have been ascribed to Mas37 (refs 4, 11–15). These results suggest that the TOM complex, which can transport all kinds of mitochondrial precursor proteins, is not sufficient for the correct integration of outer membrane proteins with a complicated topology, and instead transfers precursor proteins to the SAM complex.

The TOM machinery is the central entry gate for mitochondrial precursor proteins and consists of three receptors, Tom20, Tom22 and Tom70, the channel Tom40, and three small proteins^{2,4,5,10}. As all Tom proteins are encoded by nuclear genes, the Tom precursors themselves are also imported by means of the TOM complex. The

biogenesis of Tom40 involves nearly all the subunits of the TOM complex^{6,7,10}. Here, we addressed whether the assembly pathway of Tom40 required further mitochondrial proteins, and we found an involvement of Mas37, an outer membrane protein with a relative molecular mass of about 37,000 (M_r 37K). Pleiotropic effects have been reported for *mas37* mutants of *Saccharomyces cerevisiae*, including defects in phospholipid metabolism¹¹, protein import and genetic interaction with Tom receptors^{11,12,15}, leading to the assignment of a protein import receptor function and the name Tom37. A later study indicated that Mas37 neither was a subunit of the TOM machinery nor acted as an import receptor¹³, and thus its function remained unclear⁴.

³⁵S-labelled precursor of Tom40 was imported into isolated yeast mitochondria. On lysis with digitonin, protein complexes were

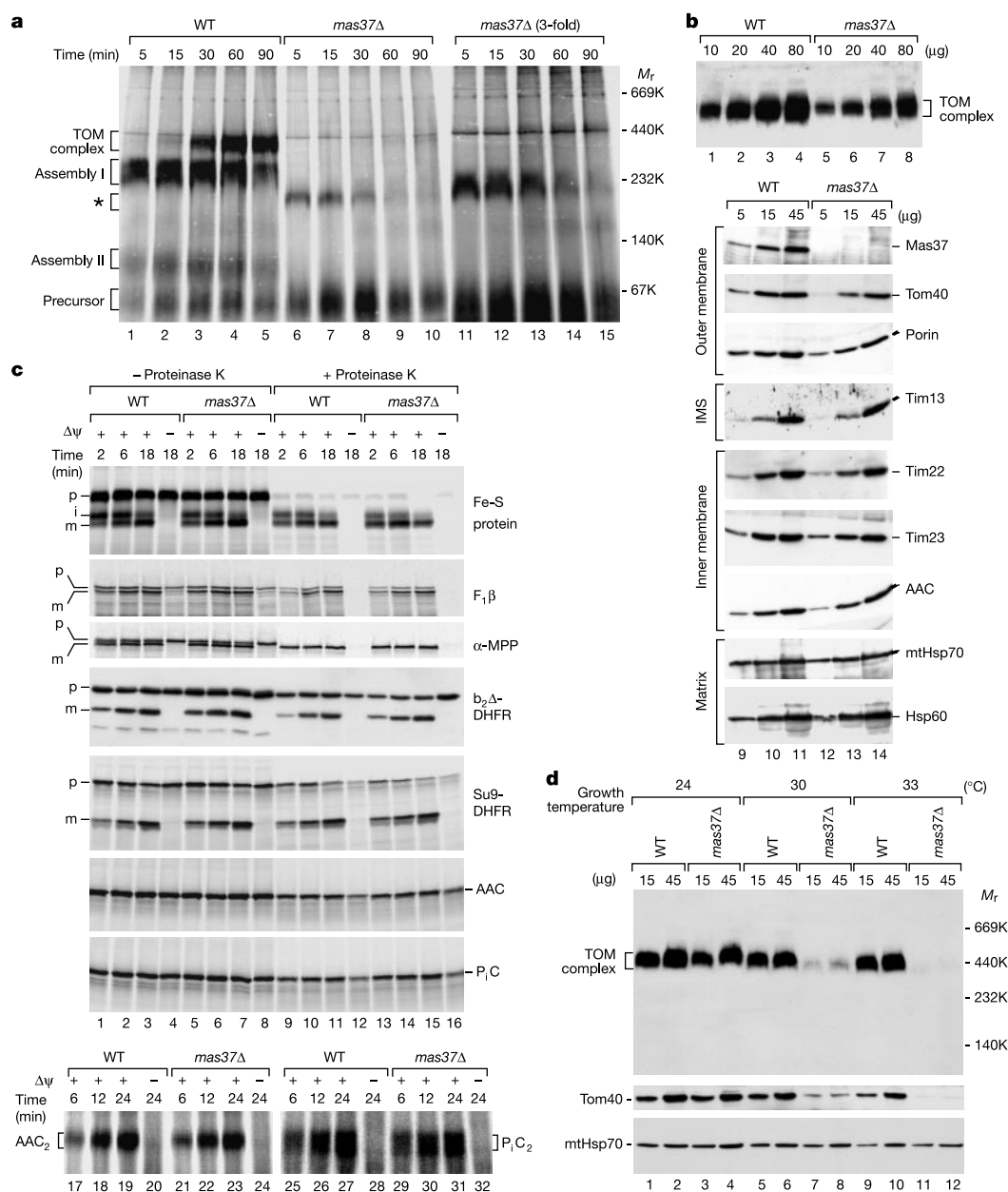


Figure 1 Mas37-deficient mitochondria are defective in biogenesis of Tom40, but not inner membrane or matrix proteins. **a**, Assembly pathway of Tom40. [³⁵S]Tom40 was incubated with mitochondria for the indicated times and analysed by BN-PAGE. WT, wild type; asterisk, 210K form in *mas37Δ*. **b**, Immunodecoration of BN-PAGE (anti-Tom40; top) or SDS-PAGE (bottom). **c**, Import of inner membrane and matrix proteins into *mas37Δ*

mitochondria. ³⁵S-labelled precursors were imported and analysed by SDS-PAGE (lanes 1–16) or BN-PAGE (lanes 17–32). $b_2\Delta$ -DHFR, purified protein¹⁶; MPP, matrix-processing peptidase; P_iC_2 , dimeric phosphate carrier; p, precursor; i, intermediate; m, mature. **d**, Reduced levels of TOM complex after growth of *mas37Δ* cells at higher temperature.

analysed by blue native polyacrylamide gel electrophoresis (BN-PAGE)⁷. In wild-type mitochondria, the assembly of Tom40 into the mature TOM core complex of 400K occurs through two intermediate complexes of about 250K and about 100K (Fig. 1a, lanes 1–5)^{6,7,9,10}. Mitochondria lacking Mas37 were strongly inhibited in the formation of all three complexes (Fig. 1a, lanes 6–15). The steady-state levels of TOM complex and pre-existing Tom40 were moderately reduced in *mas37Δ* mitochondria (Fig. 1b). The level of the outer membrane protein porin was mildly reduced. The levels of marker proteins of the other mitochondrial subcompartments did not differ between *mas37Δ* mitochondria and wild-type mitochondria (Fig. 1b). The import of precursors directed to the inner membrane or matrix was not affected, including the presequence-containing proteins Rieske Fe-S protein, F₁-ATPase subunit-β (F₁β), α-subunit of the mitochondrial processing peptidase, and two proteins containing amino-terminal targeting signals fused to dihydrofolate reductase (*b₂Δ*-DHFR and Su9-DHFR)¹⁶, as

well as the non-cleavable precursors of ADP/ATP carrier (AAC) and phosphate carrier (Fig. 1c). *mas37Δ* cells show growth defects at 30 °C and are unable to grow at 37 °C (data not shown)¹¹. *mas37Δ* cells grown at a temperature of 30 °C or higher were strongly reduced in the mitochondrial amount of TOM complex and Tom40 (Fig. 1d), leading to pleiotropic effects that explain the reported defects in protein import to the inner membrane¹¹. For all other experiments using *mas37Δ* mitochondria, the cells were grown at 23 °C to minimize indirect effects.

Wild-type mitochondria were incubated with antibodies directed against Mas37. The formation of Tom40 assembly intermediates I and II, and mature TOM complex, was blocked (Fig. 2a, lanes 7–9). Anti-Mas37 antibodies did not affect the import pathways to the matrix (F₁β) and inner membrane (AAC) (Fig. 2a). Anti-Tom5 antibodies impaired the formation of the assembly intermediate I as well as the import of F₁β and AAC (Fig. 2a)^{5,7,17}. Additionally, anti-Tom5 antibodies blocked the formation of Tom40 assembly

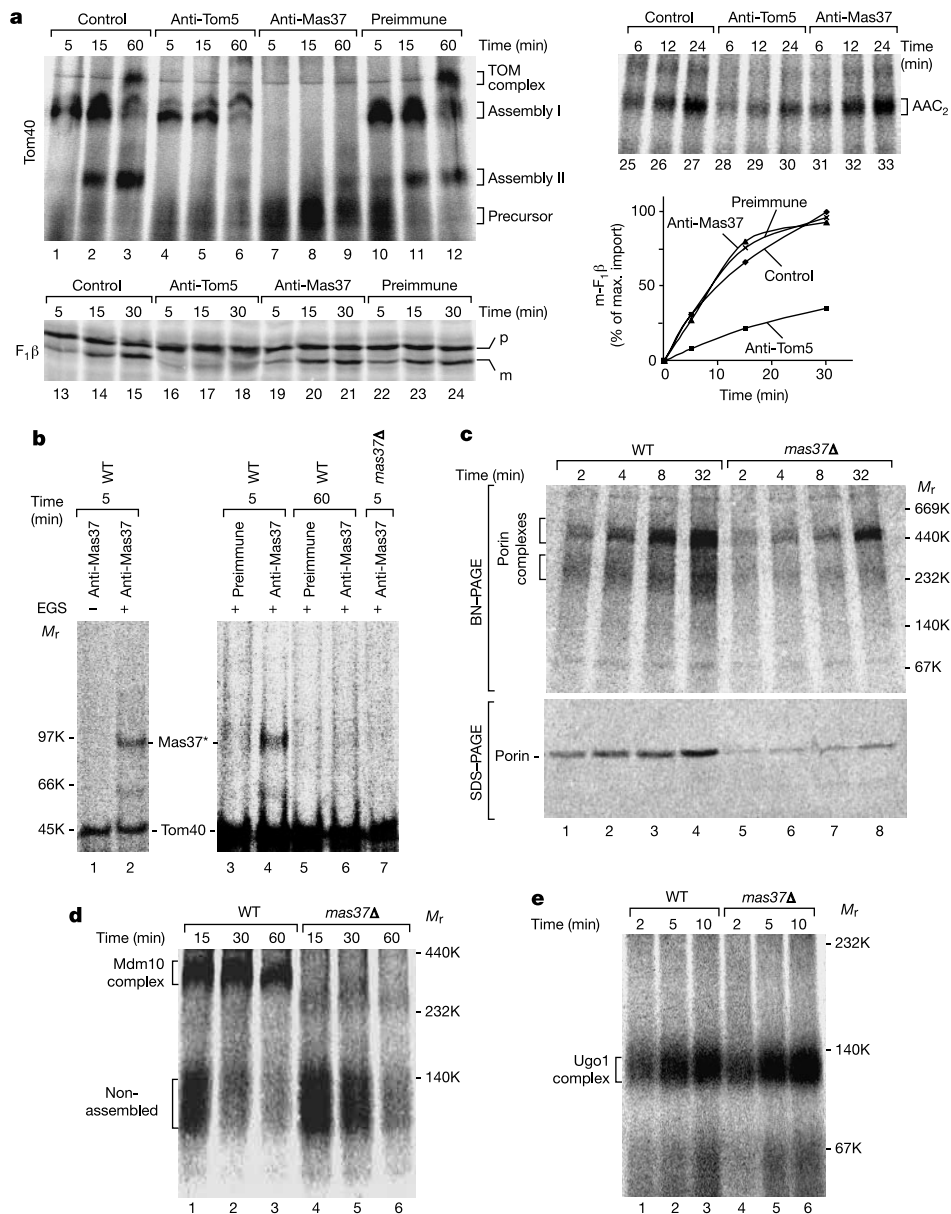


Figure 2 Mas37 is required for assembly of outer membrane proteins. **a**, Antibodies against Mas37 inhibit Tom40 assembly. Wild-type mitochondria were incubated with purified IgG, washed and incubated with ³⁵S-labelled precursors. Control, no IgG; p, precursor; m, mature. **b**, Cross-linking of Tom40 to Mas37. Mitochondria with imported

[³⁵S]Tom40 were incubated with EGS and subjected to immunoprecipitation. Asterisk, cross-linking product. **c**, Inhibition of porin assembly in *mas37Δ* mitochondria. The bottom panel includes protease treatment. **d**, Inhibition of Mdm10 assembly in *mas37Δ* mitochondria. **e**, Assembly of Ugo1 is not impaired in *mas37Δ* mitochondria.

intermediate II and mature TOM complex (Fig. 2a, lanes 4–6). We performed cross-linking with ethylene-glycolbis(succinimidylsuccinate) (EGS). On a short-term import of [35 S]Tom40 to generate the assembly intermediate I, a cross-linking product was precipitated with anti-Mas37 antibodies (Fig. 2b). The cross-linking product was neither observed in *mas37* Δ mitochondria nor after a long import into wild-type mitochondria (Fig. 2b). We conclude that Mas37 is required for the biogenesis pathway of Tom40, but not for the main protein import routes to the inner membrane.

We imported porin and Mdm10, integral membrane proteins with predicted β -strands like Tom40 (refs 18, 19). The assembly of porin⁸ was strongly impaired in *mas37* Δ mitochondria (Fig. 2c). Mdm10 was assembled into an approximately 350K complex in wild-type mitochondria, whereas in *mas37* Δ mitochondria,

formation of the 350K complex was strongly inhibited (Fig. 2d). Ugo1, an outer membrane protein containing a predicted single-membrane-spanning segment²⁰, assembled into a roughly 140K complex, yet was not affected in *mas37* Δ mitochondria (Fig. 2e). Thus, Mas37 is not required for the biogenesis of this 'simple' outer membrane protein, whereas the assembly pathway of the β -barrel proteins strongly depends on Mas37.

To address whether Mas37 is a component of the assembly I complex itself, we performed a short-term import of [35 S]Tom40. The mitochondria were incubated with specific antibodies and subjected to BN-PAGE (Fig. 3a). Anti-Mas37 antibodies as well as anti-Tom40 antibodies quantitatively shifted the assembly I complex to a high molecular mass region. Antibodies against other Tom proteins, porin or AAC did not shift the assembly I complex. As a

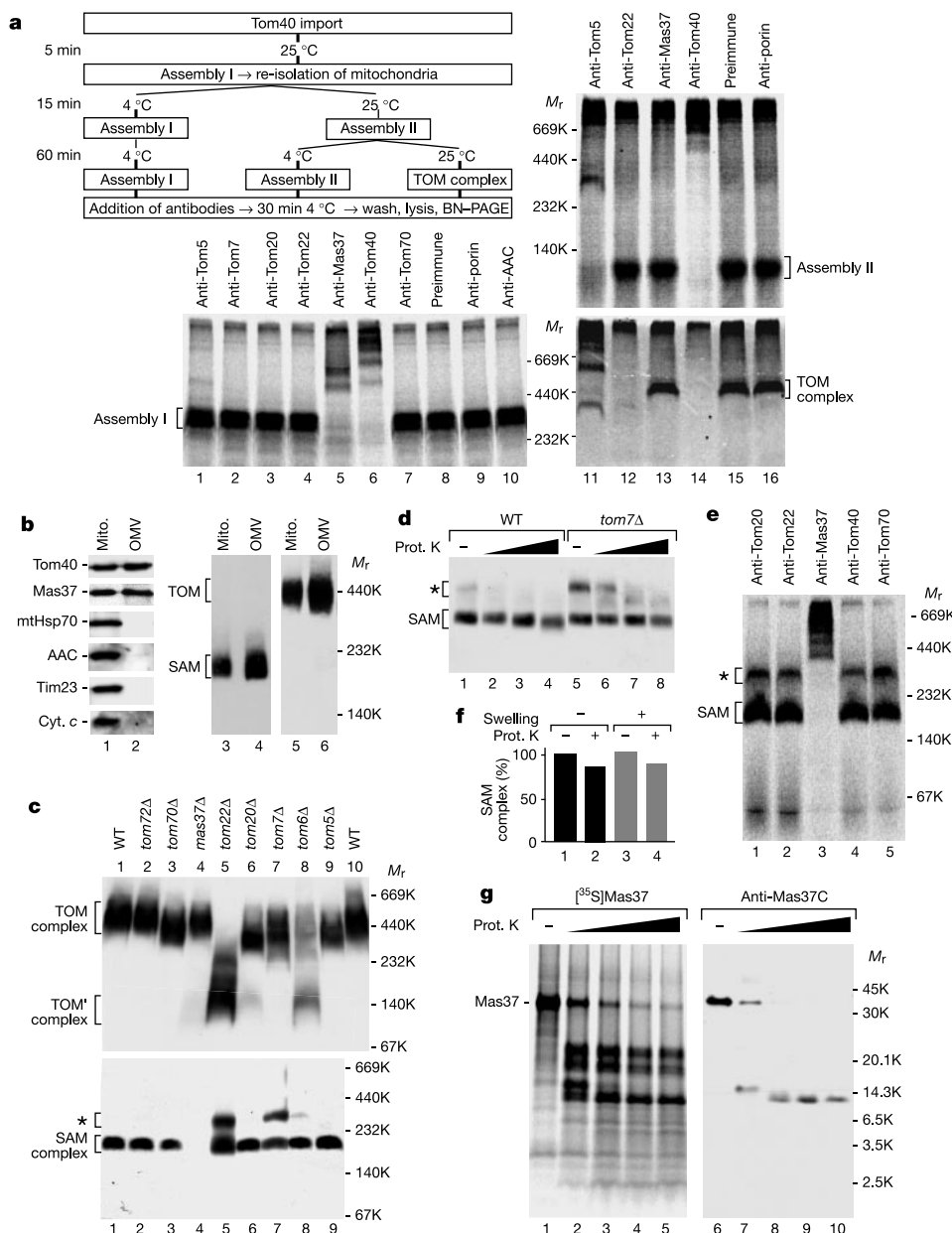


Figure 3 The SAM complex. **a**, Mas37 is present in the assembly I complex. Wild type mitochondria with accumulated [35 S]Tom40 analysed by antibody-shift BN-PAGE. **b**, SAM complex of the outer membrane. Lanes 3 and 4, anti-Mas37; lanes 5 and 6, anti-Tom40. **c**, SAM complex in *tom* mutant mitochondria. Asterisk, larger SAM complex. **d**, Sensitivity of SAM* to proteinase K (0.1–10 μ g ml⁻¹). Immunodecoration with anti-

Mas37. **e**, Antibody-shift BN-PAGE of *mas37* Δ mitochondria with imported [35 S]Mas37. **f**, Protease treatment of mitochondria and mitoplasts with 1 μ g ml⁻¹ proteinase K. **g**, Proteolytic cleavage of Mas37 (1–15 μ g ml⁻¹ proteinase K). Left, *mas37* Δ with imported [35 S]Mas37; right, wild type.

control, the assembly II complex and the mature TOM complex were shifted by anti-Tom40 antibodies, but not anti-Mas37 antibodies. Anti-Tom22 antibodies shifted the TOM complex, yet not the assembly II complex, as expected^{7,21}. Notably, anti-Tom5 antibodies not only shifted the TOM complex, but also shifted efficiently the assembly II complex. We conclude that Mas37 is a stoichiometric component of the assembly I complex, but not of the assembly II complex nor the TOM complex.

In the absence of Tom40 precursor, Mas37 was found in a roughly 210K complex of wild-type mitochondria and purified outer membranes distinct from the TOM complex (Fig. 3b). The size of 210K suggests that the 250K assembly I complex consists of the Mas37 complex and one precursor molecule of Tom40. Owing to the involvement of the Mas37 complex in the sorting and assembly pathway of outer membrane proteins, we term it the SAM complex. Mutant mitochondria lacking any of the Tom receptors or small Tom proteins still contained the SAM complex (Fig. 3c), indicating

that none of these Tom proteins is a stoichiometric subunit of the SAM complex. The TOM complex is not altered in *mas37Δ* mitochondria (Figs 1b and 3c), and a mass spectrometric analysis did not detect Mas37 in the purified TOM complex²², confirming that Mas37 is not present in the TOM complex. The SAM complex is present in the outer membrane in a molar ratio of approximately 1:4 to the TOM complex (not shown). A larger Mas37 complex (SAM*) was observed in some *tom* mutant mitochondria and also in wild-type mitochondria at lower concentration (Fig. 3c, d). Protease treatment of mitochondria removed this SAM* form (Fig. 3d), demonstrating that it contained a protease-sensitive component. Antibody-shift BN-PAGE revealed that none of the protease-sensitive Tom receptors was a stoichiometric component of the SAM or SAM* complexes (Fig. 3e), indicating that another surface protein is present in SAM*. The SAM complex itself was resistant to the protease treatment in mitochondria and mitoplasts (Fig. 3d, f). Protease treatment generated two types of Mas37 fragment (about

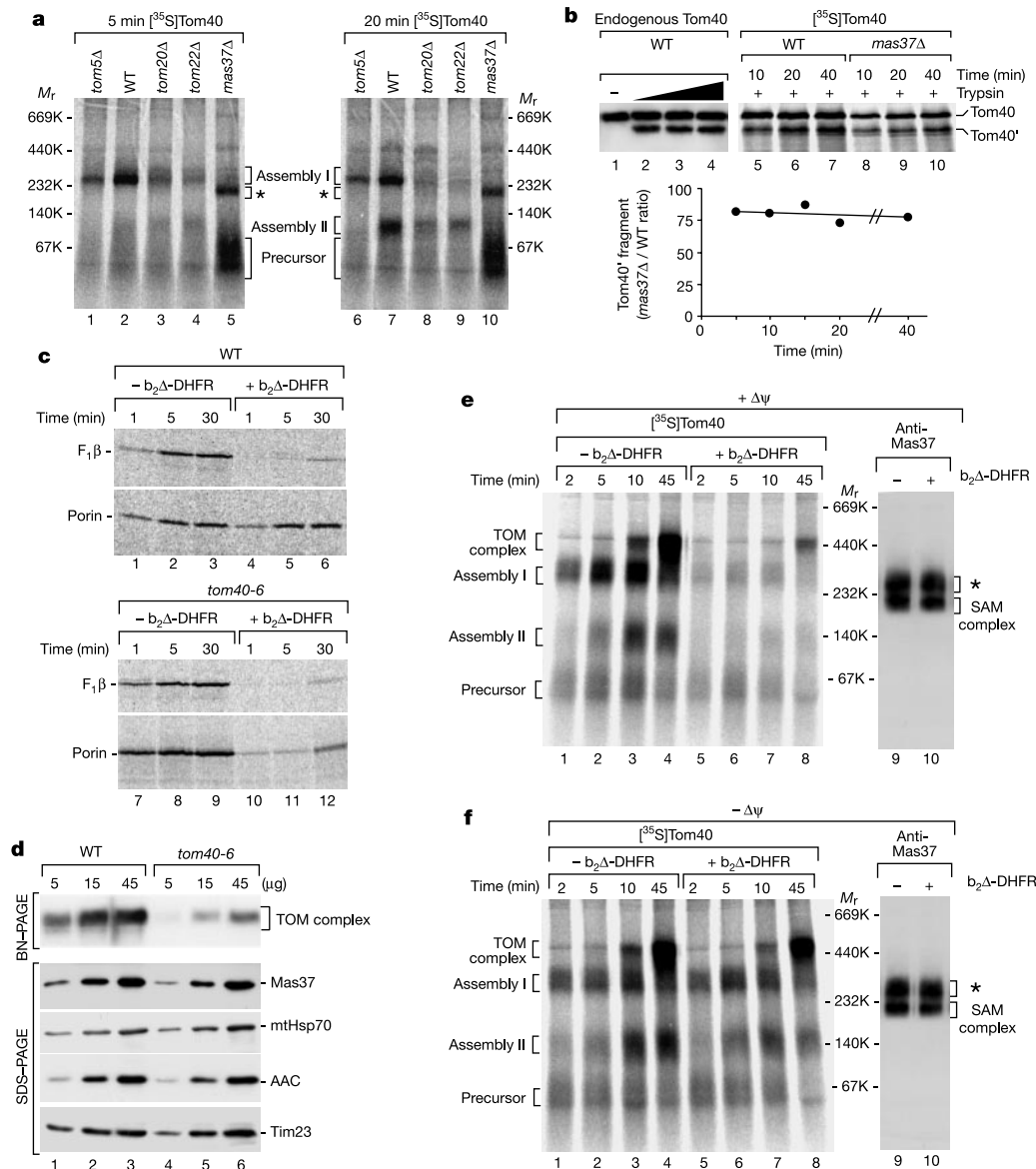


Figure 4 Tom40 precursor is transported to the SAM complex by means of the TOM complex. **a**, *tom* mutant mitochondria are impaired in formation of assembly intermediate I, yet do not accumulate non-assembled Tom40. **b**, Tom40' fragment generated by trypsin (10–30 $\mu\text{g ml}^{-1}$). **c**, Accumulation of two-membrane-spanning $b_2\Delta$ -DHFR/methotrexate¹⁶, followed by import of ³⁵S-labelled precursors and protease treatment.

d, Protein levels of *tom40-6* mitochondria. **e**, Inhibition of Tom40 assembly by accumulated $b_2\Delta$ -DHFR/methotrexate in energized *tom40-6* mitochondria. **f**, Assembly of Tom40 in *tom40-6* mitochondria after addition of $b_2\Delta$ -DHFR/methotrexate in the absence of a $\Delta\psi$.

20K and about 14K) that remained associated with mitochondria. The 14K fragment was recognized by an antibody against a carboxy-terminal region of Mas37 (Fig. 3g). The approximately 20K fragments probably represent the N-terminal portion of Mas37. Thus Mas37 exposes a protease-accessible site to the cytosol and is split into two portions by protease, yet it is not further degraded, explaining the largely unaltered mobility of the SAM complex on BN-PAGE.

The Tom40 precursor in the assembly I complex is exposed to the intermembrane space^{7,23}. As the antibody-shift BN-PAGE revealed an accessibility of the assembly intermediate I to anti-Tom40 antibodies in intact mitochondria (Fig. 3a), the accumulated Tom40 precursor probably traverses the outer membrane. Two alternative explanations were excluded. First, mitochondria loaded with anti-Tom40 antibodies were mixed with mitochondria carrying [³⁵S]Tom40 in the assembly intermediate I. On lysis, the assembly I complex was not shifted by anti-Tom40 antibodies (not shown), demonstrating that the interaction between anti-Tom40 antibodies and the assembly I complex occurred before lysis of mitochondria. Second, the SAM complex does not contain endogenous, mature Tom40 as a stoichiometric component, as anti-Tom40 antibodies did not shift the SAM complex in the absence of Tom40 precursor (Fig. 3e).

Tom20, Tom22 and Tom5 are involved in targeting of the Tom40 precursor^{6,7}. Mitochondria lacking one of these Tom proteins are strongly impaired in the formation of the assembly I complex (Fig. 4a)⁷, whereas mitochondria lacking Mas37 accumulate [³⁵S]Tom40 in a 210K complex (Figs 1a and 4a) that may represent a crippled form of the assembly I complex. Moreover, *mas37Δ* mitochondria as well as wild-type mitochondria loaded with anti-Mas37 antibodies accumulate non-assembled Tom40 precursor (Figs 1a, 2a and 4a). In contrast, *tom20Δ*, *tom22Δ* or *tom5Δ* mitochondria do not accumulate this precursor form (Fig. 4a), indicating that the TOM machinery is needed for the initial import of Tom40 before the action of the SAM complex. In wild-type mitochondria, a treatment with trypsin generates a fragment of pre-existing Tom40 (Fig. 4b, lanes 2–4)²⁴. [³⁵S]Tom40 imported *in vitro* into wild-type mitochondria as well as *mas37Δ* mitochondria gave rise to the similar fragment (Fig. 4b, lanes 5–10). The yield of formation of the Tom40' fragment in *mas37Δ* mitochondria was only mildly reduced. Thus *mas37Δ* mitochondria are strongly blocked in the formation of TOM assembly complexes, but are only mildly impaired in the initial import of Tom40. Our results also suggest a dual role for Tom5. A previous co-immunoprecipitation of Tom40 precursor with anti-Tom5 antibodies after short import times suggested a presence of Tom5 in the assembly I complex⁷. However, the SAM complex is not altered in *tom5Δ* mitochondria (Fig. 3c) and the antibody-shift BN-PAGE demonstrates that Tom5 is a stoichiometric component of the assembly II complex, but not of the assembly I complex (Fig. 3a). Tom5 is thus needed for both an early step that leads to formation of the assembly I complex, and a later step as a subunit of the assembly II complex.

We asked whether the initial import of [³⁵S]Tom40 required the general import pore (GIP) formed by endogenous Tom40 itself²⁴. We accumulated saturating amounts of b₂Δ-DHFR across both mitochondrial membranes in the presence of a membrane potential (Δψ) by stabilizing DHFR with methotrexate¹⁶. The mitochondria were re-isolated and incubated with radiolabelled precursors. The import of F₁β, but not porin, was strongly inhibited (Fig. 4c, lanes 4–6). As the TOM complexes are four times more abundant than the inner membrane translocase complexes¹⁶, only a small fraction of TOM channels are blocked in wild-type mitochondria. We used a mutant, *tom40-6*, that contained significantly lower levels of TOM complex while the levels of marker proteins, including Mas37, were not affected (Fig. 4d). The TOM complex remained functional, as F₁β, porin and Tom40 were efficiently imported (Fig. 4c, lanes 7–9, e, lanes 1–4). When b₂Δ-DHFR was accumulated in energized *tom40-6* mitochondria, the import of porin was impaired (Fig. 4c,

lanes 10–12) and the formation of the Tom40 assembly intermediate I was strongly inhibited (Fig. 4e, lanes 5–8). The assembly of [³⁵S]Tom40 was only mildly affected when mitochondria were incubated with b₂Δ-DHFR without a Δψ (Fig. 4f). As a stable arrest of b₂Δ-DHFR in the GIP requires the concomitant Δψ-dependent insertion into the inner membrane¹⁶, the two-membrane-spanning b₂Δ-DHFR was mainly responsible for the block of Tom40 import. The SAM complex itself was not affected by the accumulation of b₂Δ-DHFR (Fig. 4e, f, lane 10). We conclude that Tom40 first uses the TOM machinery, including the GIP, before it associates with the SAM complex. A transient cooperation of TOM and SAM in the transfer of precursors may also explain the observed genetic interaction between Mas37 and Tom receptors¹¹.

We have identified a protein complex in the mitochondrial outer membrane that is distinct from the TOM complex, yet has a critical role in the sorting and assembly pathway of several β-barrel outer membrane proteins. The SAM complex takes over the precursor proteins from the TOM complex and initiates the assembly process. Our findings change the current view of how proteins are sorted in an organellar membrane. It had been assumed that the translocases not only mediate the translocation of precursor proteins, but are also responsible for the proper sorting and assembly of the proteins^{2,10}. However, the sorting and assembly pathway of mitochondrial outer membrane proteins with more complicated topology, such as β-barrel proteins, is not simply mediated by the translocase (TOM complex) itself, but requires additional machinery, such that two large membrane protein complexes (TOM and SAM) function in a consecutive manner. □

Methods

In vitro import and assembly of precursor proteins

Isolated yeast mitochondria⁸ were incubated with [³⁵S]-labelled precursor proteins (rabbit reticulocyte lysate) for various times at 25 °C in import buffer (3% (w/v) bovine serum albumin (BSA), 250 mM sucrose, 5 mM MgCl₂, 80 mM KCl and 10 mM MOPS/KOH, pH 7.2) containing 2–10 mM ATP and 2 mM NADH²². For disruption of the outer membrane by osmotic swelling, mitochondria were incubated in EM buffer (10 mM MOPS-KOH, pH 7.2, 0.1 mM EDTA) for 15 min on ice. Where indicated, mitochondria were treated with 0.1–50 μg ml⁻¹ proteinase K for 15 min on ice. After washing, the re-isolated mitochondria were separated by SDS-PAGE and radiolabelled proteins were detected by digital autoradiography. For analysing the assembly of precursor proteins into complexes, mitochondria were lysed in digitonin buffer (1% (w/v) digitonin, 20 mM Tris-HCl, pH 7.4, 0.1 mM EDTA, 50 mM NaCl, 10% (w/v) glycerol, 1 mM phenylmethylsulphonyl fluoride) for 15 min on ice. BN-PAGE sample buffer (5% (w/v) Coomassie brilliant blue G-250, 100 mM Bis-Tris, pH 7.0, 500 mM 6-aminocaproic acid) was added and the extract was separated on a 6–16.5% polyacrylamide gradient gel at 4 °C^{7,16}. For cross-linking, mitochondria with imported precursors were incubated with 1 mM EGS for 30 min on ice and subjected to immunoprecipitation under denaturing conditions¹⁷. Outer membrane vesicles (OMV) were isolated as described²². In some experiments non-relevant gel lanes were excised and loading quantities of mitochondria were adjusted. Endogenous protein complexes were detected by immunodecoration and ECL detection (Amersham). Where indicated, the total amount of mitochondrial protein is shown in micrograms. Quantification was performed with ImageQuant (Molecular Dynamics).

Antibody-shift BN-PAGE of protein complexes

After *in vitro* import of [³⁵S]-labelled precursor proteins, mitochondria (20 μg protein) were resuspended in 100 μl SEM buffer (250 mM sucrose, 10 mM MOPS/KOH, pH 7.2, 0.1 mM EDTA) and incubated with purified immunoglobulin-γ (IgG) in 40 μl import buffer for 30 min on ice. After re-isolation and washing with SEM, the mitochondria were solubilized in digitonin buffer and loaded on BN-PAGE. For antibody blocking experiments, mitochondria were incubated with IgG before the import reaction.

Received 19 April; accepted 7 May 2003; doi:10.1038/nature01753.

- Koehler, C. M., Merchant, S. & Schatz, G. How membrane proteins travel across the mitochondrial intermembrane space. *Trends Biochem. Sci.* **24**, 428–432 (1999).
- Herrmann, J. M. & Neupert, W. Protein transport into mitochondria. *Curr. Opin. Microbiol.* **3**, 210–214 (2000).
- Jensen, R. E. & Johnson, A. E. Opening the door to mitochondrial protein import. *Nature Struct. Biol.* **8**, 1008–1010 (2001).
- Endo, T. & Kohda, D. Functions of outer membrane receptors in mitochondrial protein import. *Biochim. Biophys. Acta* **1592**, 3–14 (2002).
- Pfanner, N. & Geissler, A. Versatility of the mitochondrial protein import machinery. *Nature Rev. Mol. Cell. Biol.* **2**, 339–349 (2001).
- Rapaport, D. & Neupert, W. Biogenesis of Tom40, core component of the TOM complex of mitochondria. *J. Cell Biol.* **146**, 321–331 (1999).
- Model, K. *et al.* Multistep assembly of the protein import channel of the mitochondrial outer

- membrane. *Nature Struct. Biol.* **8**, 361–370 (2001).
8. Krimmer, T. *et al.* Biogenesis of porin of the outer mitochondrial membrane involves an import pathway via receptors and the general import pore of the TOM complex. *J. Cell Biol.* **152**, 289–300 (2001).
 9. Taylor, R. D., McHale, B. J. & Nargang, F. E. Characterization of *Neurospora crassa* Tom40 deficient mutants and effect of specific mutations on Tom40 assembly. *J. Biol. Chem.* **278**, 765–775 (2003).
 10. Rapaport, D. Biogenesis of the mitochondrial TOM complex. *Trends Biochem. Sci.* **27**, 191–197 (2002).
 11. Gratzner, S. *et al.* Mas37p, a novel receptor subunit for protein import into mitochondria. *J. Cell Biol.* **129**, 25–34 (1995).
 12. Hachiya, N. *et al.* Reconstitution of the initial steps of mitochondrial protein import. *Nature* **376**, 705–709 (1995).
 13. Ryan, M. T., Müller, H. & Pfanner, N. Functional staging of ADP/ATP carrier translocation across the outer mitochondrial membrane. *J. Biol. Chem.* **274**, 20619–20627 (1999).
 14. Abdul, K. M. *et al.* Functional analysis of human metaxin in mitochondrial protein import in cultured cells and its relationship with the Tom complex. *Biochem. Biophys. Res. Commun.* **276**, 1028–1034 (2000).
 15. Lithgow, T., Junne, T., Suda, K., Gratzner, S. & Schatz, G. The mitochondrial outer membrane protein Mas22p is essential for protein import and viability of yeast. *Proc. Natl Acad. Sci. USA* **91**, 11973–11977 (1994).
 16. Dekker, P. J. T. *et al.* The Tim core complex defines the number of mitochondrial translocation contact sites and can hold arrested preproteins in the absence of matrix Hsp70-Tim44. *EMBO J.* **16**, 5408–5419 (1997).
 17. Dietmeier, K. *et al.* Tom5 functionally links mitochondrial preprotein receptors to the general import pore. *Nature* **388**, 195–200 (1997).
 18. Sogo, L. F. & Yaffe, M. P. Regulation of mitochondrial morphology and inheritance by Mdm10p, a protein of the mitochondrial outer membrane. *J. Cell Biol.* **126**, 1361–1373 (1994).
 19. Martelli, P. L., Fariselli, P., Krogh, A. & Casadio, R. A sequence-profile-based HMM for predicting and discriminating β barrel membrane proteins. *Bioinformatics* **18** (Suppl. 1), S46–S53 (2002).
 20. Sesaki, H. & Jensen, R. E. UGO1 encodes an outer membrane protein required for mitochondrial fusion. *J. Cell Biol.* **152**, 1123–1134 (2001).
 21. van Wilpe, S. *et al.* Tom22 is a multifunctional organizer of the mitochondrial preprotein translocase. *Nature* **401**, 485–489 (1999).
 22. Meisinger, C. *et al.* Protein import channel of the outer mitochondrial membrane: a highly stable Tom40-Tom22 core structure differentially interacts with preproteins, small Tom proteins, and import receptors. *Mol. Cell. Biol.* **21**, 2337–2348 (2001).
 23. Matouschek, A. & Glick, B. S. Barrelling through the outer membrane. *Nature Struct. Biol.* **8**, 284–286 (2001).
 24. Hill, K. *et al.* Tom40 forms the hydrophilic channel of the mitochondrial import pore for preproteins. *Nature* **395**, 516–521 (1998).

Acknowledgements We thank M. G. Douglas, C. K. Kassenbrock, D. Cyr and A. Strub for yeast mutants, and R. Casadio for secondary structure prediction. This work was supported by the Deutsche Forschungsgemeinschaft (C.M.), Sonderforschungsbereich 388, Max Planck Research Award/Alexander von Humboldt Foundation/BMBF, the Fonds der Chemie/BMBF (N.P.) and a long-term FEBS fellowship (A.C.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for material should be addressed to N.P. (Nikolaus.Pfanner@biochemie.uni-freiburg.de).

Dimers of the N-terminal domain of phytochrome B are functional in the nucleus

Tomonao Matsushita, Nobuyoshi Mochizuki & Akira Nagatani

Department of Botany, Graduate School of Science, Kyoto University, Sakyo-ku, Kyoto 606-8502, Japan

A plant modulates its developmental processes in response to light by several informational photoreceptors such as phytochrome. Phytochrome is a dimeric chromoprotein which regulates various aspects of plant development from seed germination to flowering¹. Upon absorption of red light, phytochrome translocates from the cytoplasm to the nucleus^{2–4}, and regulates gene expression through interaction with transcription factors such as PIF3 (refs 5–7). The phytochrome polypeptide has two domains^{1,8}: the amino-terminal photosensory domain with a chromophore and the carboxy-terminal domain which contains

signalling motifs such as a kinase domain⁹. The latter is widely believed to transduce the signal to downstream components^{1,5,8–10}. Here we show that the C-terminal domain of *Arabidopsis* phytochrome B (phyB), which is known as the most important member of the phytochrome family¹, is not directly involved in signal transduction. The N-terminal domain isolated from phyB, when dimerized and localized in the nucleus, triggered full phyB responses with much higher photosensitivity than the full-length phyB. These findings indicate that the C-terminal domain attenuates the activity of phyB rather than positively transducing the signal.

We produced transformants of the *phyB* mutant of *Arabidopsis* that expressed either the N- or C-terminal domain of phyB with the green fluorescent protein (GFP) attached, hereafter called NG and CG, respectively (Fig. 1b). Confocal microscopic observation of the transgenic plants showed that CG localized in the nucleus and formed speckles as previously reported^{3,11}, whereas NG was distributed evenly throughout the nucleus and the cytoplasm (Fig. 1a). These patterns were unaffected by exposure to light. Since phyB mediates the inhibition of the hypocotyl elongation under continuous red light¹², we measured the reduction in hypocotyl lengths in these lines under continuous red light. Unexpectedly, the plants expressing NG showed a significant response to continuous red light whereas those expressing CG did not (Fig. 1c). This indicated that NG retained signalling activity but not CG. This was not due to the effect of GFP because the N-terminal domain without GFP (N) was as active as NG (Fig. 2b, c). The present observation is not in conflict with the dominant negative activity of N reported in the wild-type background¹³. Less active N could compete with endogenous phyB for the downstream components to reduce signalling.

We made transgenic plants expressing NG in combination with a nuclear localization signal (NLS) or a nuclear export signal (NES) to determine the site of NG signal transduction within the cell (Fig. 2b). Since relatively small proteins often enter the nucleus passively, we also produced plants expressing NG fused to β -glucuronidase (GUS), with which passive entry into the nucleus can be avoided¹⁴. As expected, NG-NLS and NG-GUS-NLS accumulated in the nucleus, but did not form speckles even under light (Fig. 2a). By contrast, NG-NES and NG-GUS were localized exclusively in the cytoplasm, indicating that the N-terminal domain lacks nuclear

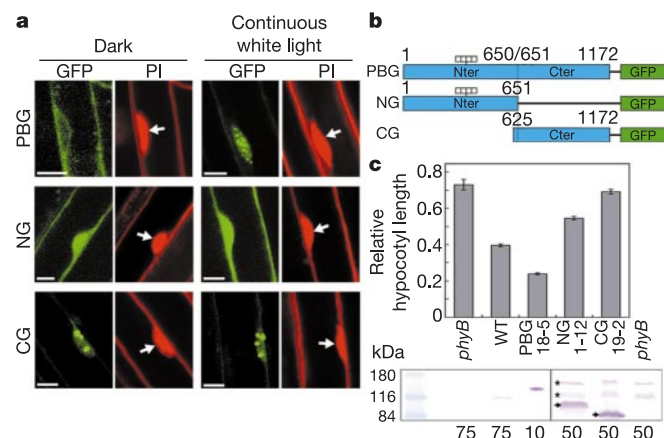


Figure 1 Subcellular localization and biological activity of the N- and C-terminal domains of phyB. **a**, Subcellular localization in dark and continuous white light. The GFP and propidium iodide (PI) images in the same view are shown. Arrows indicate nuclei. Scale bars, 10 μ m. **b**, Diagrams of phyB derivatives. The dotted line shows the junction between the N- (Nter) and C-terminal (Cter) domains. The four rectangles represent the chromophore. **c**, Hypocotyl lengths under continuous red light ($4.9 \mu\text{mol m}^{-2} \text{s}^{-1}$). WT, wild-type. The bottom panel of **c** shows immunoblotting with mAb1 (left) and anti-GFP antibody (right). The amounts of loaded total protein, in μ g, are indicated at the bottom. Two nonspecific bands (asterisks) were observed. Arrows indicate the specific bands.

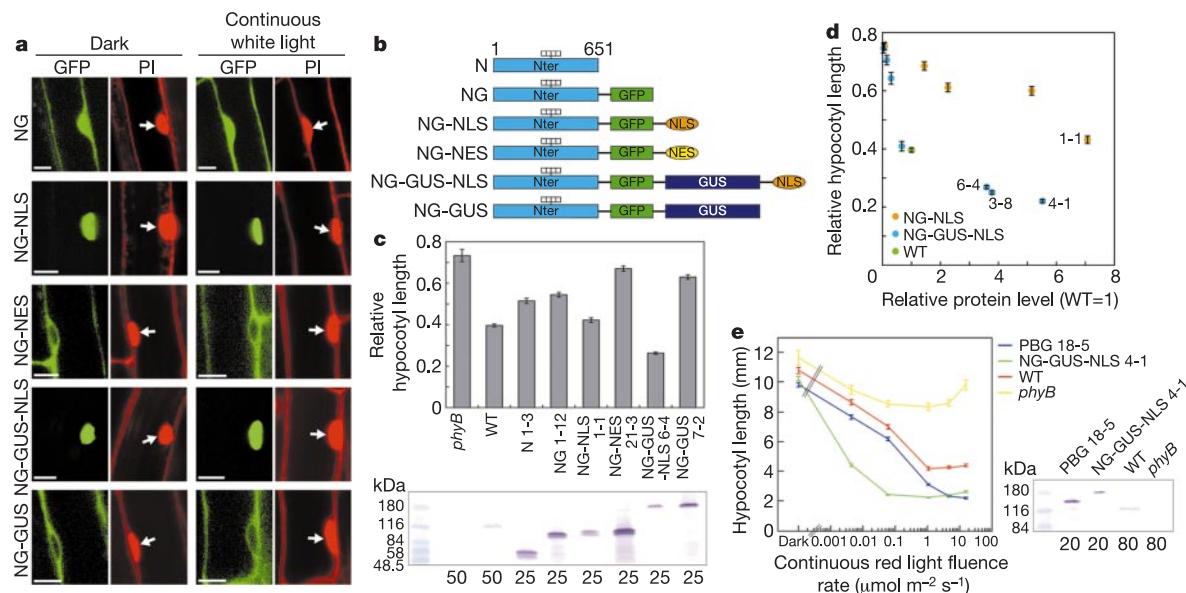


Figure 2 Signalling activity of the N-terminal domain of phyB in the nucleus. **a, b**, Subcellular localization (**a**) and diagrams (**b**) of the phyB derivatives. Details are as in Fig. 1a and b. **c**, Hypocotyl lengths under continuous red light ($4.9 \mu\text{mol m}^{-2} \text{s}^{-1}$). The bottom panel of **c** shows immunoblotting with mBA1. **d**, Hypocotyl length under

continuous red light ($4.9 \mu\text{mol m}^{-2} \text{s}^{-1}$) as a function of the introduced protein level expressed as a ratio to the endogenous phyB level in the wild type. The numbers beside the spots denote the plant line. **e**, Fluence-rate response curves for hypocotyl lengths under continuous red light. The right panel of **e** shows immunoblotting with mBA1.

localization activity. The hypocotyl assay indicated that NG-NLS and NG-GUS-NLS had higher phyB activity than either NG-NES or NG-GUS (Fig. 2c). Hence, we concluded that NG transduced the signal in the nucleus.

In the hypocotyl assay, NG-GUS-NLS was even more active than NG-NLS (Fig. 2c). The phyB activity under a saturating light condition in several independent transgenic lines that exhibited different levels of the introduced protein further confirmed this activity (Fig. 2d). The phyB-activity-enhancing effect of GUS might be due to its multimerizing activity¹⁵, because the N-terminal domain separated from the C-terminal domain exists as a monomer whereas the full-length phytochrome exists in nature as a dimer¹³. This possibility was examined by native gel electrophoresis (Supplementary Information 1). The ratios of the estimated native size to the monomeric mass were 3.19 and 3.26, respectively, for the full-

length phyB fused to GFP (PBG) and NG-GUS-NLS, which correspond to the ratio in the authentic full-length phytochrome^{13,16}. Although some GUS chimaeric proteins form tetramers as well as dimers¹⁵, we could not detect a tetrameric band of NG-GUS-NLS. By contrast, the ratio for NG-NLS was 1.57, which is half that for either PBG or NG-GUS-NLS. Hence, we concluded that NG-GUS-NLS was dimerized but NG-NLS was not, although a minor portion of NG-GUS-NLS could exist as a tetramer. In summary, the N-terminal domain can transduce the signal in the monomeric form but dimerization is required for full activity.

The hypocotyl assay was performed on plants expressing NG-GUS-NLS or PBG to measure the phyB activity under various intensities of continuous red light (Fig. 2e). NG-GUS-NLS was about a hundred times more sensitive than PBG, although the immunoblot showed slightly higher expression of the introduced protein in the PBG plant than in the NG-GUS-NLS plant (Fig. 2e). Since PBG and PBG-NLS exhibited similar photosensitivity (data not shown), the difference was not due to NLS. These results suggest that the C-terminal domain inhibits the signalling of the N-terminal domain in the full-length phyB. The assay also revealed that NG-GUS-NLS showed no activity in darkness (Fig. 2e), indicating that accumulation of the N-terminal domain in the nucleus is not sufficient for triggering the response. Consistently, the NG-GUS-NLS bearing a mutation at the chromophore binding site failed to respond to light (Supplementary Information 2). We then confirmed that NG-GUS-NLS could mediate other phyB responses, such as chlorophyll accumulation, expansion of cotyledons and inhibition of petiole elongation^{1,12} (Fig. 3a–c). In all cases, NG-GUS-NLS exhibited higher activity than PBG. The morphology of the transgenic and wild-type plants is shown in Fig. 3d.

Some missense mutations within the C-terminal domain of phytochrome reduce the biological activity of each phytochrome without affecting its spectral nature or dimerization competence¹⁰. How do these mutations affect the activity of phyB? Plants expressing PBG with the most severe G767-to-R mutation, PBG(G767/R), did not show phyB activity (Fig. 4b, c), although this mutation did not affect the dimerization of PBG¹⁰ (data not shown). PBG(G767/R) was localized mainly in the cytoplasm even under light (Fig. 4a). This implied that PBG(G767/R) was less active because it could not

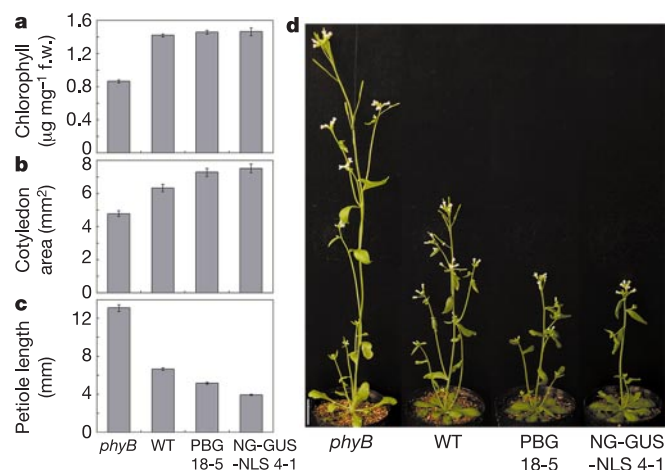


Figure 3 NG-GUS-NLS can trigger various phyB responses. **a, b**, Chlorophyll levels (**a**) and cotyledon area (**b**) of seedlings grown for one week under continuous white light ($50 \mu\text{mol m}^{-2} \text{s}^{-1}$). f.w., fresh weight. **c**, Petiole lengths in plants grown for three weeks under continuous white light ($44 \mu\text{mol m}^{-2} \text{s}^{-1}$). **d**, Mature plants grown for four weeks under continuous white light ($80 \mu\text{mol m}^{-2} \text{s}^{-1}$). Scale bar, 2 cm.

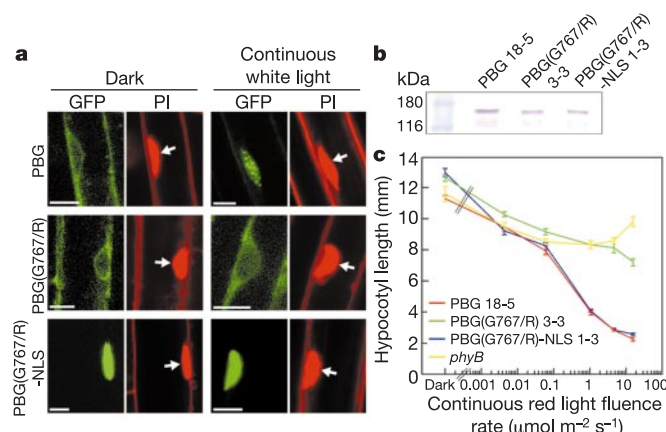


Figure 4 The G767/R mutation impairs accumulation of phyB to nucleus. **a**, Subcellular localization of the phyB derivatives. Details are as in Fig. 1a. **b**, Immunoblotting of the phyB derivatives with mAb1. Each lane contained 25 μg total protein. **c**, Fluence-rate response curves for hypocotyl lengths under continuous red light.

accumulate in the nucleus. To examine this possibility further, we made transgenic plants expressing PBG(G767/R) with NLS attached, PBG(G767/R)-NLS. This fusion protein was accumulated in the nucleus as expected, but did not form speckles (Fig. 4a). The hypocotyl assay clearly indicated that PBG(G767/R)-NLS was as active as PBG (Fig. 4b, c). This observation was further confirmed for other phyB-regulated responses (data not shown). Hence, we concluded that the G767/R mutation lowered the phyB activity by hindering accumulation of the protein in the nucleus, although PBG(G767/R) is capable of transducing the signal when localized in the nucleus.

On the basis of our present findings, we propose that the N-terminal domain, but not the C-terminal domain, of phyB transduces the signal to downstream components (Fig. 5). This view is consistent with the previous observation that the N-terminal domain determines the mode of phytochrome action^{17,18}. The speckle formation has been proposed to be required for phytochrome signal transduction⁴. This may not be true because phyB derivatives such as NG-GUS-NLS (Fig. 2) and PBG(G767/R)-NLS (Fig. 4) triggered the phyB responses without forming speckles.

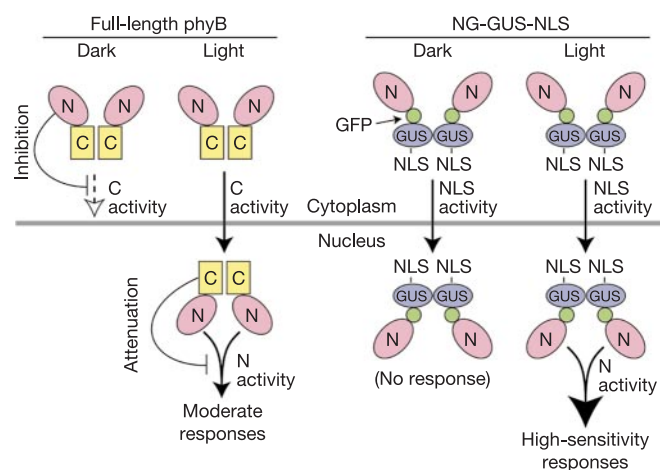


Figure 5 Model of the action of phyB and NG-GUS-NLS in the cell. PhyB and NG-GUS-NLS enter the nucleus by the activities of the C-terminal domain of phyB (C) and NLS, respectively. In darkness, the N-terminal domain (N) acts to block the entry of phyB into the nucleus. Nuclear accumulation of NG-GUS-NLS without light is not sufficient to elicit the response. In the light, dimerized N of phyB and NG-GUS-NLS transduces the signal efficiently. The signalling activity of N is attenuated by C in phyB.

At present, it remains unclear how the N-terminal domain transduces the signal in the nucleus. It may interact with transcription factors such as PIF3, originally said to interact with the C-terminal domain of phytochrome³, but later shown to interact with the N-terminal domain of phyB in a light-dependent manner^{6,19}. The C-terminal domain appears to attenuate the activity of phyB (Fig. 2e). The kinase activity found in this domain⁹ may be involved in this function. In fact, the phosphorylation of several serine residues in the N-terminal domain of phyA has been suggested to serve as a desensitizing mechanism^{8,20}. The factors such as ELF3 (ref. 21), ADO1/ZTL/LKP1 (refs 22, 23) and PKS1 (ref. 24), which interact with the C-terminal domain of phyB, affect the physiology of phyB. These factors may modify the signalling activity of the N-terminal domain of phyB through the interaction with the C-terminal domain. Rapid phytochrome responses, such as changes in intracellular Ca^{2+} levels²⁵ are beyond the scope of the present analysis. Such responses could be triggered by a totally different mechanism. □

Methods

Plant materials, growth conditions and measurements

Landsberg erecta (Ler) ecotype *Arabidopsis thaliana* plants with a null allele *phyB-5* mutation¹² were used. Seedlings were grown at 22 °C on 0.6% agar plates containing Murashige-Skoog (MS) medium, supplemented with 2% (w/v) sucrose unless stated otherwise. For the hypocotyl assay, seedlings were grown on MS agar plates without sucrose for five days. Relative hypocotyl lengths were expressed as a ratio to the respective dark value. For immunoblot analysis, seedlings were grown for one week under continuous white light ($44 \mu\text{mol m}^{-2} \text{s}^{-1}$). The chlorophyll assay was performed as previously described²⁶. Petiole lengths were measured after the longest leaf from each plant was photographed. The white and red light sources used were described previously². Data represent the mean $\pm$ s.d. ($n = 25$ –30) for each measurement.

Plasmid construction

SV40 NLS²⁷ was obtained by annealing two oligonucleotides (5'-GCCTAAGAAGAAG AGAAAGGTTGAGGATAGCCCGGCTGCA-3' and 5'-GCCCGGGCTATCCTCCAAC CTTTCTCTTCTTCTTAGGCTGCA-3'), encoding the amino acid residues LQPKKKRKVGG(stop). NES from PKI²⁸ was generated by annealing the following oligonucleotides: 5'-GAACGAGCTTGCTCTTAAGTTGGCTGGACTTGATAT TAACAAGACT GGAGGATAGCCCGGCTGCA-3' and 5'-GCCCGGGCTATCCT CCACTCT TGTTAATATCAAGTCCAGCCAACCTAAGAGCAAGCTCGTCTGCA-3', encoding the amino acid residues LQNELALKLAGLDINKTGG(stop). The fragments encoding the N- and C-terminal domains of phyB were amplified by PCR and the *PHYB* moiety of PBG² was replaced to obtain NG and CG, respectively. They were fused to the sequences encoding NLS or NES and/or PCR-amplified *GUS* fragments. The G767/R and C357/A mutations were generated by using the QuikChange Site-Directed Mutagenesis Kit (Stratagene). All the constructs were inserted between the constitutive cauliflower mosaic virus 35S promoter and the Nos terminator of pPZP211/35S-nosT²⁶.

Plant transformation and regeneration of transgenic lines

The *phyB-5* mutant was transformed with the plasmids described above using the *Agrobacterium*-mediated floral dip method²⁹ and transformants were selected on the MS medium containing 25 mg ml^{-1} kanamycin and 166 mg ml^{-1} claforan (Hoechst). For each construct, at least 20 independent lines that segregated approximately 3:1 for kanamycin-resistance in the T2 generation were established. For physiological experiments, a few representative lines were selected on the basis of the level of overexpression. For all experiments, T3 self-progeny of homozygous T2 plants were used. The kanamycin resistance co-segregated tightly with the GFP fluorescence.

Analysis of subcellular localization in transgenic seedlings

Transgenic *Arabidopsis* seedlings grown in darkness or under continuous white light of $44 \mu\text{mol m}^{-2} \text{s}^{-1}$ were stained with 20 $\mu\text{g ml}^{-1}$ propidium iodide (PI) (Molecular Probes) to visualize nuclei and cell walls. The subcellular localization of the GFP and PI fluorescence was visualized using a confocal laser scanning microscope (Zeiss LSM510) with the FITC channel (green, GFP) and the TRITC channel (red, PI). For etiolated seedlings, a dim green safelight was used in every step before microscopy and photographs were taken during the first minute of observation. At least 20 independent transgenic lines were analysed for each construct. The observations were done in every part of the seedlings and the same results were obtained in each line. The results obtained with root cells are therefore shown as representative. All constructs analysed in this study were also expressed transiently in onion cells by particle bombardment and found to exhibit the same subcellular localization (data not shown).

Immunochemical experiments

Most of the procedures were performed as described previously². Direct protein extraction using hot SDS buffer was also performed and similar results were obtained for all the plants analysed in this study (data not shown). Protein levels were measured by

densitometric analysis of immunoblots. The monoclonal antibody mBA1 is specific to the N-terminal domain of phyB³⁰. Anti-GFP monoclonal antibody was from Sigma.

Received 26 February; accepted 19 May 2003; doi:10.1038/nature01837.

1. Smith, H. Phytochromes and light signal perception by plants—an emerging synthesis. *Nature* **407**, 585–591 (2000).
2. Yamaguchi, R., Nakamura, M., Mochizuki, N., Kay, S. A. & Nagatani, A. Light-dependent translocation of a phytochrome B:GFP fusion protein to the nucleus in transgenic *Arabidopsis*. *J. Cell Biol.* **145**, 437–445 (1999).
3. Nagy, F., Kircher, S. & Schäfer, E. Nucleo-cytoplasmic partitioning of plant photoreceptors phytochromes. *Semin. Cell Dev. Biol.* **11**, 505–510 (2000).
4. Kircher, S. *et al.* Nucleocytoplasmic partitioning of the plant photoreceptors phytochrome A, B, C, D, and E is regulated differentially by light and exhibits a diurnal rhythm. *Plant Cell* **14**, 1541–1555 (2002).
5. Ni, M., Tepperman, J. M. & Quail, P. H. PIF3, a phytochrome-interacting factor necessary for normal photoinduced signal transduction, is a novel basic helix-loop-helix protein. *Cell* **95**, 657–667 (1998).
6. Ni, M., Tepperman, J. M. & Quail, P. H. Binding of phytochrome B to its nuclear signalling partner PIF3 is reversibly induced by light. *Nature* **400**, 781–784 (1999).
7. Martínez-García, J. F., Huq, E. & Quail, P. H. Direct targeting of light signals to a promoter element-bound transcription factor. *Science* **288**, 859–863 (2000).
8. Park, C.-M., Bhoo, S.-H. & Song, P.-S. Inter-domain crosstalk in the phytochrome molecules. *Semin. Cell Dev. Biol.* **11**, 449–456 (2000).
9. Yeh, K.-C. & Lagarias, J. C. Eukaryotic phytochromes: light-regulated serine/threonine protein kinases with histidine kinase ancestry. *Proc. Natl Acad. Sci. USA* **95**, 13976–13981 (1998).
10. Wagner, D. & Quail, P. H. Mutational analysis of phytochrome B identifies a small COOH-terminal domain region critical for regulatory activity. *Proc. Natl Acad. Sci. USA* **92**, 8596–8600 (1995).
11. Sakamoto, K. & Nagatani, A. Nuclear localization activity of phytochrome B. *Plant J.* **10**, 859–868 (1996).
12. Reed, J. W., Nagpal, P., Poole, D. S., Furiya, M. & Chory, J. Mutations in the gene for the red/far-red light receptor phytochrome B alter cell elongation and physiological responses throughout *Arabidopsis* development. *Plant Cell* **5**, 147–157 (1993).
13. Wagner, D., Koloszar, M. & Quail, P. H. Two small spatially distinct regions of Phytochrome B are required for efficient signaling rates. *Plant Cell* **8**, 859–871 (1996).
14. Grebenok, R. J. *et al.* Green-fluorescent protein fusions for efficient characterization of nuclear targeting. *Plant J.* **11**, 573–586 (1997).
15. Kato, A., Hayashi, M. & Nishimura, M. Oligomeric proteins containing N-terminal targeting signals are imported into peroxisomes in transgenic *Arabidopsis*. *Plant Cell Physiol.* **40**, 586–591 (1999).
16. Cherry, J. R. *et al.* Carboxy-terminal deletion analysis of oat phytochrome A reveals the presence of separate domains required for structure and biological activity. *Plant Cell* **5**, 565–575 (1993).
17. Wagner, D., Fairchild, C. D., Kuhn, R. M. & Quail, P. H. Chromophore-bearing NH₂-terminal domains of phytochromes A and B determine their photosensory specificity and differential light lability. *Proc. Natl Acad. Sci. USA* **93**, 4011–4015 (1996).
18. Clough, R. C. *et al.* Sequences within both the N- and C-terminal domains of phytochrome A are required for PFR ubiquitination and degradation. *Plant J.* **17**, 155–167 (1999).
19. Shimizu-Sato, S., Huq, E., Tepperman, J. M. & Quail, P. H. A light-switchable gene promoter system. *Nature Biotechnol.* **10**, 1041–1044 (2002).
20. Stockhaus, J. *et al.* Serine-to-alanine substitutions at the amino-terminal region of phytochrome A result in an increase in biological activity. *Genes Dev.* **6**, 2364–2372 (1992).
21. Liu, X. L., Covington, M. F., Fankhauser, C., Chory, J. & Wagner, D. R. *ELF3* encodes a circadian clock-regulated nuclear protein that functions in an *Arabidopsis* *PHYB* signal transduction pathway. *Plant Cell* **13**, 1293–1304 (2001).
22. Somers, D. E., Schultz, T. F., Milnamow, M. & Kay, S. A. *ZEITLUPE* encodes a novel clock-associated PAS protein from *Arabidopsis*. *Cell* **101**, 319–329 (2000).
23. Jarillo, J. A. *et al.* An *Arabidopsis* circadian clock component interacts with both *CRY1* and *phyB*. *Nature* **410**, 487–490 (2001).
24. Fankhauser, C. *et al.* PK1, a substrate phosphorylated by phytochrome that modulates light signaling in *Arabidopsis*. *Science* **284**, 1539–1541 (1999).
25. Shacklock, P. S., Read, N. D. & Trewavas, A. J. Cytosolic free calcium mediates red light-induced photomorphogenesis. *Nature* **358**, 753–755 (1992).
26. Mochizuki, N., Brusslan, J. A., Larkin, R., Nagatani, A. & Chory, J. *Arabidopsis* genomes uncoupled 5 (*GUN5*) mutant reveals the involvement of Mg-chelatase H subunit in plastid-to-nucleus signal transduction. *Proc. Natl Acad. Sci. USA* **98**, 2053–2058 (2001).
27. Kalderon, D., Roberts, B. L., Richardson, W. D. & Smith, A. E. A short amino acid sequence able to specify nuclear location. *Cell* **39**, 499–509 (1984).
28. Wen, W., Meinke, J. L., Tsien, R. Y. & Taylor, S. S. Identification of a signal for rapid export of proteins from the nucleus. *Cell* **82**, 463–473 (1995).
29. Clough, S. J. & Bent, A. F. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16**, 735–743 (1998).
30. Shinomura, T. *et al.* Action spectra for phytochrome A- and B-specific photoinduction of seed germination in *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA* **93**, 8129–8133 (1996).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank S. Kaihara, R. W. Ridge and T. Takeyama for critical reading of the manuscript; R. Tsugeki for providing *GUS* clone; and laboratory members for support and discussion. This work was supported by grants from the Program for Promotion of Basic Research Activities for Innovative Biosciences and from Special Coordination Funds for Promoting Science and Technology from the Science and Technology Agency and by Grants-in-Aid for Scientific Research (B) and for Scientific Research on Priority Areas from the Ministry of Education, Science, Sports and Culture of Japan.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.N. (nagatani@physiol.bot.kyoto-u.ac.jp).

Processivity of the single-headed kinesin KIF1A through biased binding to tubulin

Yasushi Okada¹, Hideo Higuchi² & Nobutaka Hirokawa¹

¹Department of Cell Biology and Anatomy, University of Tokyo, Graduate School of Medicine, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

²Department of Metallurgy, Graduate School of Engineering, and Center for Interdisciplinary Research, Tohoku University, Sendai 980-8579, Japan

Conventional isoforms of the motor protein kinesin behave functionally not as ‘single molecules’ but as ‘two molecules’ paired. This dimeric structure poses a barrier to solving its mechanism^{1–4}. To overcome this problem, we used an unconventional kinesin KIF1A (refs 5, 6) as a model molecule. KIF1A moves processively as an independent monomer^{7,8}, and can also work synergistically as a functional dimer⁹. Here we show, by measuring its movement with an optical trapping system¹⁰, that a single ATP hydrolysis triggers a single stepping movement of a single KIF1A monomer. The step size is distributed stochastically around multiples of 8 nm with a gaussian-like envelope and a standard deviation of 15 nm. On average, the step is directional to the microtubule’s plus-end against a load force of up to 0.15 pN. As the source for this directional movement, we show that KIF1A moves to the microtubule’s plus-end by ~3 nm on average on binding to the microtubule, presumably by preferential binding to tubulin on the plus-end side. We propose a simple physical formulation to explain the movement of KIF1A.

As the tag for the manipulation and measurement, we attached a 0.2-μm bead to one end of the wild-type KIF1A construct A382, residues 1–382 of mouse KIF1A (not containing the neck of conventional kinesin^{7,8}; Supplementary Information 1). The rapid response time (0.25 ms)¹⁰ and the small pivotal and the rotational brownian movement of the 0.2-μm bead enabled us to detect clearly the small and fast movement at low load¹¹. For most assays, biotinylated A382 was attached to the streptavidin-coated bead. For some assays, non-biotinylated A382 was attached to the bead via the Fab fragment of an anti-His-tag antibody to avoid the possibility of pseudo-dimerization on streptavidin. Furthermore, to examine the contribution by the structural change at the carboxy-terminal neck linker¹², the bead was attached to the amino terminus or the C terminus of A382 (Fig. 1a).

When in contact with a microtubule on a coverglass, the KIF1A bead showed stochastic oscillation (Fig. 1b). The movement was not affected by the immobilization method of KIF1A or by its anchoring position on the bead. The Poisson statistics analysis of the fraction of the moving beads as a function of the mixing ratio of KIF1A to beads^{13,14} confirmed that a single KIF1A molecule was sufficient for the movement (Supplementary Information 2). For the experiments below, we used a mixing ratio at which the fraction of the bead movement was less than 0.3, so that more than 99% of the moving beads would be driven by only one molecule¹⁰. At 10-fold higher mixing ratios, some of the beads showed qualitatively different movement (Supplementary Information 2), similar to the movement by artificially dimerized KIF1A (ref. 9). The lack of this type of movement confirmed that we were actually analysing single-motor events.

The stochastic oscillation of the KIF1A bead cannot be explained by simple brownian movement interrupted by binding events, because the oscillation shifted to the microtubule plus-end, far exceeding the range of thermal fluctuation (Fig. 1b; Supplementary Information 3). Furthermore, the same KIF1A bead moved processively to the microtubule plus-end for more than 1 μm when the

optical trap was turned off (Fig. 1c). This indicates that the oscillation reflects the active tug-of-war between the movement of KIF1A and the external load by the fixed-beam optical trap. With conventional kinesin, this tug-of-war results in a stall, because it rarely moves backwards¹¹. In contrast, KIF1A sometimes moves backwards even under no load^{7,8}. External load would affect the probability of forward movement, leading to stochastic oscillation. Consistently, the mean velocities at various force levels were linearly related to the force (Fig. 1d).

When ATPase turnover was slowed by decreasing the ATP concentration, the KIF1A bead showed a clear bidirectional stepwise movement (Fig. 2a, b). The interval between steps (dwell time) and the duration of the stepping movement (stepping time) followed single-exponential distributions (Fig. 2c, d). The mean of the dwell time was equal to the duration of the nucleotide-free state of a single ATPase cycle (Fig. 2e), indicating that a single ATP hydrolysis produces a single mechanical step (Supplementary Information

3). The size and the direction of this step varied stochastically (Fig. 2f), a behaviour very different from the regular 8-nm step of a conventional kinesin dimer^{11,15}. However, the step size was distributed around multiples of 8 nm (Fig. 2f, g), reflecting the periodicity of the binding site on the microtubule^{12,16}. The envelope of this discrete distribution was approximately gaussian. Its standard deviation (15 ± 2 nm) was insensitive to load, whereas its mean showed a linear dependence on load (Fig. 2h, i).

This gaussian-type envelope of the distribution of the step size indicates that the stepping movement contains one or more diffusion-like processes, which agrees with our previous result⁷. However, diffusion alone cannot produce a directional movement. Directional movement should occur either immediately before diffusion (Fig. 3c, step 2) or immediately after diffusion (Fig. 3c, step 4). To examine these possibilities, we measured the movement at the transition from the ADP-bound diffusion state into the rigor-binding state by using the experimental framework established to detect the movement of myosin when it binds to actin¹⁷. The position of the KIF1A bead was measured before and after its binding to the microtubule (Fig. 3a). To confirm that KIF1A bound correctly to the microtubule, ATP was added to the buffer ($2 \mu\text{M}$) and only the binding events followed by the stepping movement were scored. In this condition, free KIF1A remains in the ADP-bound state owing to the very slow release of ADP (less than 0.003 s^{-1}). Then KIF1A releases ADP and binds to one of the binding sites. From the transient kinetic measurements of other kinesins^{4,18}, this transition state is expected to have the same properties as the intermediate state during processive movement. The position before and after the binding to the microtubule was measured by averaging out the pivotal brownian movement of the bead.

The ensemble average of $\sim 2,000$ binding events (Fig. 3b) showed a rapid plus-end-directed displacement after binding, followed by a slow plus-end-directed movement. The latter slow movement agreed with the mean velocity of the KIF1A movement, reflecting the ensemble average of KIF1A's stepping motion by ATP hydrolysis (Fig. 3a, blue arrow). The size of the rapid displacement was estimated by extrapolating the slope of the slow movement to zero time as 2.8 ± 0.8 nm. This burst-type displacement is coupled with the release of ADP or the binding of KIF1A to the microtubule before ATP binding, because ATP binding takes ~ 250 ms on average at $2 \mu\text{M}$ ATP. The simplest interpretations of the rapid displacement are that KIF1A preferentially selects the binding site on the plus-end side or changes the structure by ADP release as proposed by ATP binding^{3,12}.

On the basis of these results, we can quantitatively model the stepping movement of KIF1A (Fig. 3c, Supplementary Information 5). The hydrolysis of newly bound ATP into ADP releases KIF1A from the binding site on the microtubule (Fig. 3c, step 2). Let the (putative) directional movement of the KIF1A bead during this process be δ_{release} (in nm). This parameter has not yet been measured, but contains the displacement by the power stroke and other putative mechanical processes. The ADP-bound form of KIF1A moves along the microtubule by one-dimensional diffusion (step 3) with diffusion coefficient D ($26 \pm 6 \text{ nm}^2 \text{ ms}^{-1}$; Supplementary Information 1). The theory of brownian movement holds that the mean and variance of the displacement under load force F are given by $-(\tau D/k_B T)F$ and $2D\tau$, respectively (Supplementary Information 5), where k_B , T and τ denote the Boltzman constant, the temperature (299 ± 1 K) and the duration of this diffusive movement (4 ± 1 ms, measured as the stepping time in Fig. 2c), respectively. Finally, KIF1A binds again to the microtubule and ADP is released (Fig. 3c, step 4). KIF1A moves to the plus-end by δ_{bind} (2.8 ± 0.8 nm) on average (Fig. 3b). The step movement is the sum of these factors, and its mean and standard deviation are calculated to be

$$\langle s \rangle = \delta_{\text{release}} + \delta_{\text{bind}} - (D/k_B T)\tau F = \delta_{\text{release}} + 2.8 - 25F \text{ (in nm)}$$

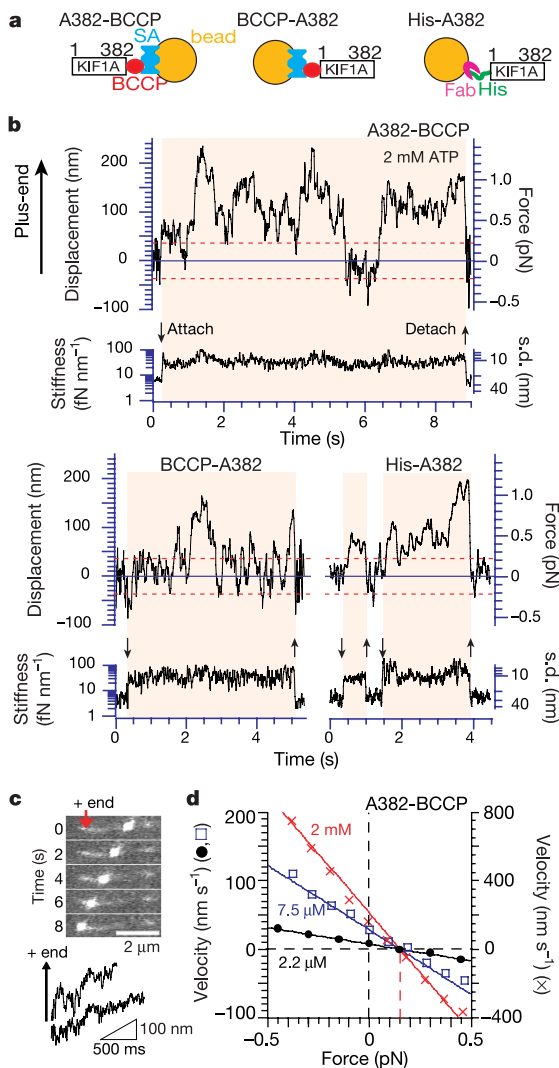


Figure 1 Movement of KIF1A beads. **a**, Design of the bead. BCCP, biotinylation tag; SA, streptavidin; Fab, Fab fragment of anti-His-tag antibody (His). (See Supplementary Information 1 for details.) **b**, Displacement of KIF1A beads. Red lines show the range of the thermal fluctuation (s.d.) of the free bead. Traces of stiffness indicate the attachment and detachment of the KIF1A to and from the microtubule. **c**, Processive movement of the KIF1A bead after the optical trap has been turned off. **d**, Force-velocity relationship. The lines show the predictions from the model in Fig. 3c and Supplementary Information 5. Crosses, 2 mM ATP; squares, 7.5 μM ATP; circles, 2.2 μM ATP.

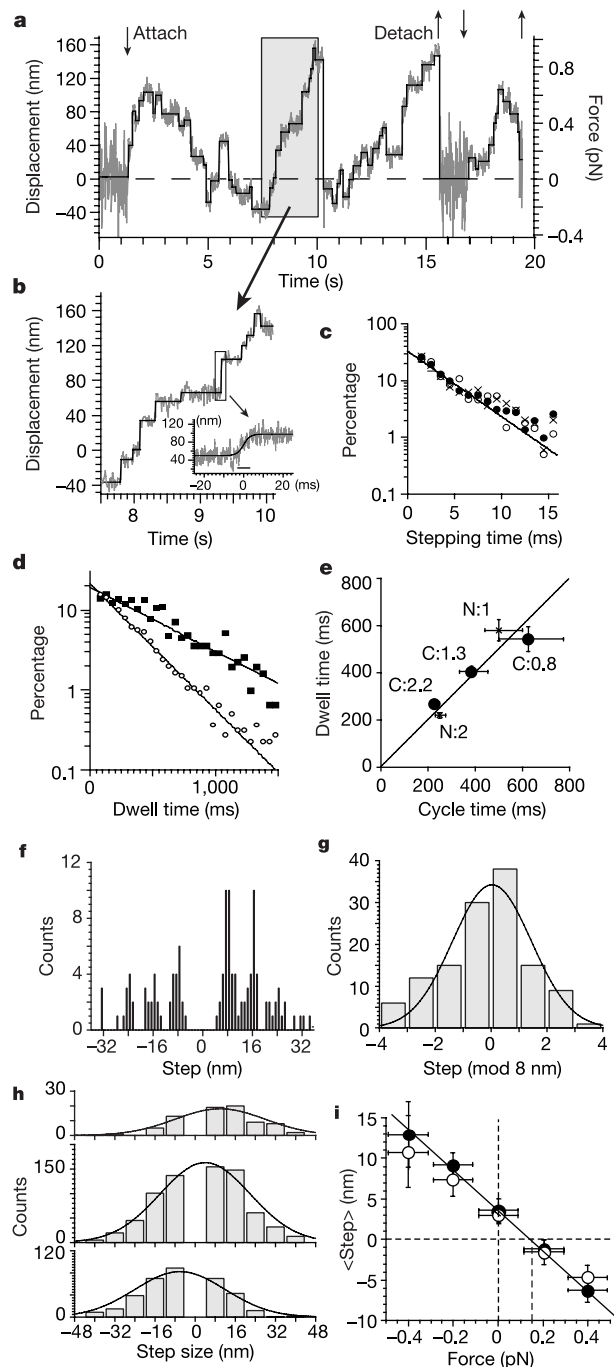


Figure 2 Stepping movement of KIF1A beads. **a**, Raw (grey) and filtered (black) displacement (A382-BCCP; [ATP] = 2.2 μ M). **b**, Higher temporal resolution view of **a**. **c**, ATP-independent (squares, 0.8 μ M; crosses, 1.3 μ M; circles, 2.2 μ M) distribution of the stepping time. **d**, ATP-dependent (squares, 0.8 μ M; circles, 2.2 μ M) distribution of the dwell time. In **d**, squares, 0.8 μ M ATP; circles, 2.2 μ M ATP. **e**, Comparison of the dwell time (**d**) with the ATPase cycle time (Supplementary Information 1). The construct (N, BCCP-A382; C, A382-BCCP) and the ATP concentration in μ M are indicated next to the data points. **f**, Distribution of the step size (A382-BCCP). **g**, Deviation of the step size around multiples of 8 nm. The data in **f** are re-plotted with the best-fit gaussian curve (mean 0, s.d. 3 nm). **h**, Load dependence of the distribution of the step size (A382-BCCP, 2.2 μ M ATP). Steps of at least ± 8 nm were scored for their size and the force level, and plotted with the best-fit gaussian curves (see Methods). A positive load means that the bead is pulled towards the microtubule minus end. Top panel, -0.5 to -0.3 pN; middle panel, -0.1 to 0.1 pN; bottom panel, 0.3 to 0.5 pN. **i**, Linear dependence of the mean step on load, at 2 μ M ATP. The line shows the theoretical prediction on the basis of $\langle s \rangle = \delta - (D/k_B T)F$. Filled circles, A382-BCCP; open circles, BCCP-A382. (See the inset to Supplementary Information 5 for details.)

and

$$\sigma(s) = (2D\tau)^{1/2} = 14 \text{ (in nm)}$$

respectively. These theoretical estimates fit the observations quantitatively (Fig. 2h, i), and the value for δ_{release} is estimated to be 0.8 ± 1 nm (Fig. 2i). The same theoretical results also explain the linear force-velocity relationship (Fig. 1d). The velocity of the movement V is expressed as $V = \langle s \rangle k_{\text{cat}} = (3.6 - 25F)k_{\text{cat}}$. Under the assumption that the load force does not significantly affect the turnover rate of ATP hydrolysis, k_{cat} , in this weak load range, this equation gives a linear force-velocity relationship (Supplementary Information 5). This equation with the k_{cat} values for A382 (Supplementary Information 1) agrees well with observation (Fig. 1d).

The nearly zero value for the estimate of δ_{release} implies that the

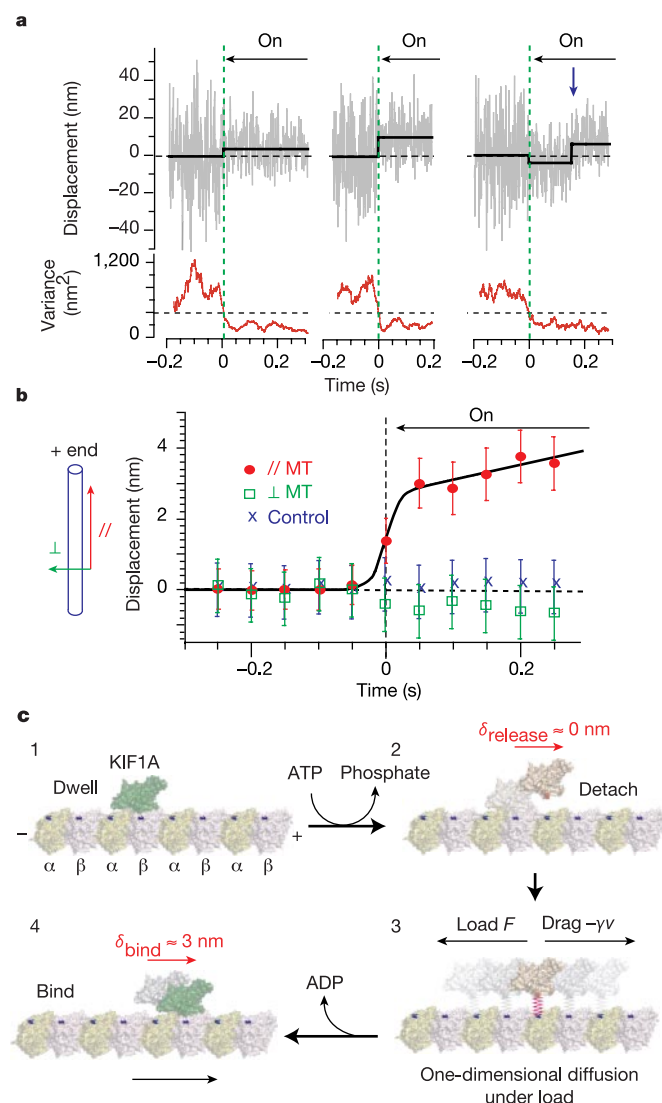


Figure 3 Plus-end-directed movement after binding to the microtubule. **a**, Raw (grey) and filtered (black) displacement with variance (red). The blue arrow shows the stepping by the ATPase reaction. **b**, Ensemble mean $\pm$ s.e.m. of the KIF1A bead displacement (red) after binding to the microtubule. To minimize possible artefacts in the shift, about half of the measurement ($n = 875$) was with plus-end-up microtubules (on the monitor) and the rest ($n = 774$) was with plus-end-down microtubules. Means of the bead position 50 ms after binding to the plus-end-up and plus-end-down microtubules were 3.2 and 3.1 nm, respectively, indicating no artefact in the shift. As controls for the null net displacement, the off-axis displacement (green, $n = 1649$) and the displacement after non-specific binding to the coverglass (blue, $n = 1420$) are also indicated. (See Methods for details.) **c**, Summary of results.

power stroke or the putative movement during the period from the ATP binding to its hydrolysis into ADP does not contribute much to the net movement. This is consistent with our failure to detect a difference in the step size between N-terminally and C-terminally anchored beads (Fig. 2i), because the power stroke of the C-terminal neck would be reflected in a difference in the movement between beads anchored at the different termini. This also implies that the movement of KIF1A can be modelled simply by the theory of a flush-ratchet-type brownian motor^{19,20}. In this model, δ_{bind} reflects the bias in selection of the binding site. This bias might have been introduced by the artificial linkage of the motor to the bead. However, a free KIF1A motor moves processively, like the motor attached to the bead (Supplementary Information 1). Thus, the effect of anchoring to the bead is negligible and the bias reflects the asymmetry in the binding potential between KIF1A and the microtubule^{7,8}.

As predicted from the theories of the brownian motor, the efficiency of the monomeric KIF1A movement was very low. The mechanical force and work by a single KIF1A are only 0.15 pN and $\sim 0.1k_B T$, respectively, $\sim 0.4\%$ of the free-energy change of ATP hydrolysis ($\sim 25k_B T$). With the clustering of a few KIF1A molecules on the bead surface, the bead moved at $\sim 800 \text{ nm s}^{-1}$ up to $\sim 2.5 \text{ pN}$ load force (Supplementary Information 2). This enhancement is also explained by the same theory of the brownian motor²⁰. The increase in the efficiency and force is caused by suppression of the diffusive movement of each head by the mechanical link to other heads (Supplementary Information 6). The putative chemical coordination between the two heads²¹ and the structural change at the neck linker^{3,12} further increases the efficiency and force. The observed directional binding (Fig. 3c, step 4) reflects the mechanism that enables the leading head to bind to the forward tubulin subunit but not to the backward subunit during the hand-over-hand walking of the two-headed kinesin to ensure efficient work. Furthermore, the structural similarities between kinesin, myosin and G proteins indicate that the energetic strategies and the asymmetrical binding for movement or signal transduction might be universal. This would also provide suggestions for the design principles of artificial nanomachines. $\square$

Methods

Preparation of KIF1A beads

The design, purification and basic characterization of the KIF1A proteins are described in Supplementary Information 1. Latex beads $0.2 \mu\text{m}$ in diameter were coated with streptavidin or the F(ab')₂ fragment of the anti-His-tag antibody^{22,23}. After incubation with KIF1A for 1–4 h, the bead was diluted with buffer (Supplementary Information 1) immediately before the measurement. A custom-made optical trapping apparatus¹⁰ was used to measure the movement of the KIF1A bead on the polarity-marked microtubule²⁴ attached to the coverglass. All assays were performed at 25–27 °C.

Data analysis

Bead displacements were recorded at a sampling rate of 20 kHz with a bandwidth of 10 kHz. The force of KIF1A was calculated from the bead displacement multiplied by the trap stiffness (6 fN nm^{-1}), which was determined from the variance of the thermal fluctuations of a trapped bead by the equipartition theorem of energy¹¹.

For analysis of the force-velocity relationship, the traces of the bead movement were segmented with a time window that was double the ATPase cycle time. The mean velocity and mean load force were calculated and scored for each segment.

Steps were detected as a rapid (less than 20 ms) positional change (more than 4 nm) followed by a long dwell (more than 10 ms) from the Harr wavelet-filtered trace²⁵. Pre-step and post-step positions of the bead were calculated by averaging the raw displacement data. The step size was estimated from the difference between the pre-step and post-step positions. The mean force level was determined as the load force at the middle of the step (the mean of the pre-step and post-step positions). The dwell time was determined as the interval from one step to the next. The dwell followed by the detachment or by the more than 100-nm step back to the zero position was excluded from the statistics. The stepping time was estimated by fitting the raw data to a sigmoid curve:

$x_0 + d/[1 + \exp\{-4(t - t_0)/\tau\}]$, where x_0 , d and τ denote the pre-step position, the step size and the stepping time, respectively (Fig. 2b, inset). Only steps larger than 20 nm were analysed, for accuracy in measurement of the stepping time.

For ensemble averaging, the binding events were automatically detected by thresholding the variance of the bead position (10-ms time window), and were synchronized at the decrease in the variance (green lines in Fig. 3a). Only binding events whose duration was longer than 250 ms and which were followed by the stepping movement were analysed ($n = 1,649$). The mean displacement (50-ms time window) from the averaged pre-binding position was fitted with a sigmoidal bump followed by a linear displacement: $d/[1 + \exp(-4t/\tau)] + vt$; $v = 0$ ($t < 0$), v_0 ($t \geq 0$). As a control, the non-specific binding of the bead to the coverglass in the absence of microtubules was analysed similarly ($n = 1,420$).

Received 21 February; accepted 19 May 2003; doi:10.1038/nature01804.

- Hua, W., Chung, J. & Gelles, J. Distinguishing inchworm and hand-over-hand processive kinesin movement by neck rotation measurements. *Science* **295**, 844–848 (2002).
- Kaseda, K., Higuchi, H. & Hirose, K. Coordination of kinesin's two heads studied with mutant heterodimers. *Proc. Natl Acad. Sci. USA* **99**, 16058–16063 (2002).
- Rice, S. *et al.* A structural change in the kinesin motor protein that drives motility. *Nature* **402**, 778–784 (1999).
- Gilbert, S. P., Moyer, M. L. & Johnson, K. A. Alternating site mechanism of the kinesin ATPase. *Biochemistry* **37**, 792–799 (1998).
- Okada, Y., Yamazaki, H., Sekine Aizawa, Y. & Hirokawa, N. The neuron-specific kinesin superfamily protein KIF1A is a unique monomeric motor for anterograde axonal transport of synaptic vesicle precursors. *Cell* **81**, 769–780 (1995).
- Hirokawa, N. Kinesin and dynein superfamily proteins and the mechanism of organelle transport. *Science* **279**, 519–526 (1998).
- Okada, Y. & Hirokawa, N. Mechanism of the single-headed processivity: diffusional anchoring between the K-loop of kinesin and the C terminus of tubulin. *Proc. Natl Acad. Sci. USA* **97**, 640–645 (2000).
- Okada, Y. & Hirokawa, N. A processive single-headed motor: Kinesin superfamily protein KIF1A. *Science* **283**, 1152–1157 (1999).
- Tomishige, M., Klopstein, D. R. & Vale, R. D. Conversion of Unc104/KIF1A kinesin into a processive motor after dimerization. *Science* **297**, 2263–2267 (2002).
- Nishiyama, M., Muto, E., Inoue, Y., Yanagida, T. & Higuchi, H. Substeps within the 8-nm step of the ATPase cycle of single kinesin molecules. *Nature Cell Biol.* **3**, 425–428 (2001).
- Nishiyama, M., Higuchi, H. & Yanagida, T. Chemomechanical coupling of the forward and backward steps of single kinesin molecules. *Nature Cell Biol.* **4**, 790–797 (2002).
- Kikkawa, M. *et al.* Switch-based mechanism of kinesin motors. *Nature* **411**, 439–445 (2001).
- Block, S. M., Goldstein, L. S. & Schnapp, B. J. Bead movement by single kinesin molecules studied with optical tweezers. *Nature* **348**, 348–352 (1990).
- Endow, S. A. & Higuchi, H. A mutant of the motor protein kinesin that moves in both directions on microtubules. *Nature* **406**, 913–916 (2000).
- Svoboda, K., Schmidt, C. F., Schnapp, B. J. & Block, S. M. Direct observation of kinesin stepping by optical trapping interferometry. *Nature* **365**, 721–727 (1993).
- Kikkawa, M., Okada, Y. & Hirokawa, N. 15 Å resolution model of the monomeric kinesin motor, KIF1A. *Cell* **100**, 241–252 (2000).
- Veigel, C. *et al.* The motor protein myosin-I produces its working stroke in two steps. *Nature* **398**, 530–533 (1999).
- Hackney, D. D. Pathway of ADP-stimulated ADP release and dissociation of tethered kinesin from microtubules. Implications for the extent of processivity. *Biochemistry* **41**, 4437–4446 (2002).
- Astumian, R. D. Thermodynamics and kinetics of a Brownian motor. *Science* **276**, 917–922 (1997).
- Julicher, F., Ajdari, A. & Prost, J. Modeling molecular motors. *Rev. Mod. Phys.* **69**, 1269–1281 (1997).
- Farrell, C. M., Mackey, A. T., Klumpp, L. M. & Gilbert, S. P. The role of ATP hydrolysis for kinesin processivity. *J. Biol. Chem.* **277**, 17079–17087 (2002).
- Berliner, E., Young, E. C., Anderson, K., Mahtani, H. K. & Gelles, J. Failure of a single-headed kinesin to track parallel to microtubule protofilaments. *Nature* **373**, 718–721 (1995).
- Inoue, Y. *et al.* Movements of truncated kinesin fragments with a short or an artificial flexible neck. *Proc. Natl Acad. Sci. USA* **94**, 7275–7280 (1997).
- Howard, J. & Hyman, A. A. Preparation of marked microtubules for the assay of the polarity of microtubule-based motors by fluorescence microscopy. *Methods Cell Biol.* **39**, 105–113 (1993).
- deCastro, M. J., Fondecave, R. M., Clarke, L. A., Schmidt, C. F. & Stewart, R. J. Working strokes by single molecules of the kinesin-related microtubule motor ncd. *Nature Cell Biol.* **2**, 724–729 (2000).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank T. J. Mitchison for the gift of GMPCPP; H. Fukuda, H. Sato and M. Sugaya for technical and secretarial assistance; and J. Howard, M. Kikkawa and our colleagues for discussion. This work was supported by a Center of Excellence Grant-in-Aid (N.H.) and a Grant-in-Aid (H.H.) from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to N.H. (hirokawa@m.u-tokyo.ac.jp).

The latest for the lab

Murine accomodation, proteomics and FDA compliance.



The red house: home comforts for mice.

Mouse House

Nalgene www.nalgenelabware.com
Creature comforts

The Mouse House is a triangular unit made of red polycarbonate with a flat top that fits into the corner of the cage. It is designed to enhance the well-being of laboratory mice by providing an environment that reduces abnormal stereotyped behaviours while encouraging a wide range of natural activities, such as jumping, climbing, squeezing through holes and sleeping during daylight hours. The red colour, which is seen as black by mice, provides a secure hiding place for them while being transparent to humans. The Mouse House is especially useful for immunodeficient and transgenic mice because of the heightened sensitivity and expense of these research models. The unit is washable and autoclavable.

4000 Q TRAP

Applied Biosystems www.appliedbiosystems.com
Linear ion trap mass spectrometer

The 4000 Q TRAP LC/MS/MS system, latest addition to Applied Biosystems' linear ion trap family, is a single-platform solution for the pharmaceutical drug development process. The system combines triple quadrupole technology with ion trap sensitivity, to give both qualitative and quantitative capabilities for protein and small molecule research. For

proteomics researchers, the system can help with the rapid identification and measurement of post-translational modifications at low levels from complex biological samples in a single LC/MS/MS run. In addition, it is designed to enable drug discovery and development researchers to rapidly identify, confirm and quantify metabolites at extremely low levels in a single experiment.

SPRINTPrep

Agencourt Bioscience www.agencourt.com
Gives you a run for your money

SPRINTPrep is a rapid and easily automated purification method for high-copy plasmids that can be completed in less than 15 minutes. It isolates plasmid DNA directly from overnight growth cultures without the need to harvest cells, and can shorten sequencing time by as much as a full day. The simple protocol requires to sample transfers, cell pelleting, centrifugation or filtration. It is available in 96- and 384-well formats.

Gemini 4.1/ Magellan Tracker 4.0

Tecan www.tecan.com
CFR rules OK

An increasing number of laboratories need to comply with the FDA's CFR Part 11 ruling on electronic records and signatures, which specifies additional requirements in the areas of user management, authentication, electronic signature, audit trail and data security. These rules are fully implemented in Tecan's Gemini 4.1 pipetting software and Magellan Tracker 4.0 program for instrument control and data reduction. Gemini 4.1 can be used for life science applications on the Genesis RSP, Genesis Freedom and Genesis workstations, including those enabled by its Te-MO multi-channel pipettors. Magellan Tracker 4.0 has improved on-screen user instructions (wizards) and is suitable for use on detection products ranging from Sunrise to the new Ultra Evolution.



Four into one: FAST Frame from S&S.

FAST Frame

Schleicher & Schuell BioScience
www.schleicher-schuell.de
Multi-slide plate

The FAST Frame plate is designed for high-throughput processing of microarrays, and is used with 16- pad or 8-pad FAST slides and the corresponding incubation chambers. It holds up to 4 FAST slides, allowing processing of up to 64 different protein microarrays simultaneously. The 9 mm spacing (microtitre plate format) of the array pads on FAST slides makes the loaded FAST Frame compatible with automated liquid-handling systems and 8-channel manual pipettes. The rows and columns on each plate are labelled for easy indexing and sample application. Slides and chambers are easily inserted and removed, yet held securely in place. The affixed incubation chambers create leak-free wells that allow the use of multiple samples without cross-contamination. The frame is compatible with other 25×76 mm glass slides when they are used in conjunction with Schleicher & Schuell's 8- and 16-well incubation chambers.

These notes are compiled in the Nature office from information provided by the manufacturers. Material intended for publication can be sent to newproducts@nature.com.

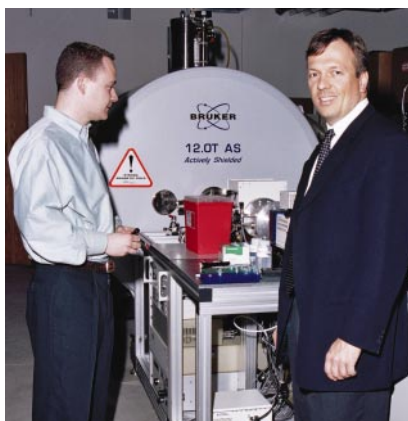
Probing the proteome

The completion of the human genome sequence, coupled with analytical techniques such as mass spectrometry, has fuelled interest in proteomics. Diane Gershon reports.

Even before the draft sequence of the human genome was completed, investors and companies were looking to the 'next big thing' — proteomics. Huge amounts of time and money were invested in proteomics-related ventures, and a crop of start-up companies emerged. Some firms set up large-scale proteomics factories running around the clock, with the aim of taking an inventory of the entire human proteome. A few years down the line, reality is setting in. How to turn proteomics data into products remains an open question and investors are becoming impatient.

"I think what it shows is that this is not a simple technology space and the barriers to entry are very high," says Keith Williams, chief executive of Proteome Systems in North Ryde, Australia. But nothing has changed in terms of what people need to do, he says. "Proteins are absolutely at the centre of where the products are going to be." Most drugs currently on the market, or in development, are directed against protein targets.

And the field still offers a major market opportunity for instrument and reagent companies. According to international analysts Frost & Sullivan, the worldwide market for proteomics technologies is projected to



Frank Laukien (right) and Michael Easterling with Bruker Daltonics' \$2-million, 12-tesla FT-mass spectrometer.

more than double between 2002 and 2005, increasing from an estimated US\$2.4 billion to \$5.8 billion.

Proteomics encompasses systematic studies of the identity, changing abundance, distribution, modifications, interactions, structure and function of large sets of proteins, as well as their involvement in disease.

But the technical challenge cannot be

overstated. Multiple proteins can be obtained from each gene, giving vastly more proteins in the human proteome than the 30,000–40,000 genes estimated for the genome. Proteins are also inherently unstable outside a narrow range of environmental conditions. And the wide and dynamic range of protein abundances makes it hard to detect low-abundance proteins against a high-abundance background.

The sheer number of different tasks and objectives in proteomics means that there is unlikely to be a universal technology "that enables you to identify all proteins, in all samples, all of the time", says Kevin Auton, chief executive of proteomics technology developers NextGen Sciences of Huntingdon, UK.

Although equipment tailor-made for proteomics is beginning to appear on the market, access to the technology, some of which comes with a high price tag, is still an issue for many researchers. More automated and integrated proteomics platforms that are both robust and user-friendly are needed, particularly for proteome-wide analyses.

Mass spectrometry (MS) has been key to the development of proteomics. It can be used to identify and, increasingly, quantify large numbers of proteins from complex samples

BRUKER DALTONICS

MASS SPECTROMETRY GOES MAINSTREAM

At the less expensive end of the market is the ion-trap mass spectrometer, which is reasonably sensitive and has a reputation for being rugged, reliable and easy to use. But conventional three-dimensional ion traps have limited trap capacity, although newer instruments offer improvements on this front. The 1100 Series LC-MSD Trap XCT non-linear ion trap from Agilent in Palo Alto, California, for example, has a trap capacity four to eight times greater than its predecessor, which the company says will give a tenfold increase in sensitivity. Prices start at about US\$200,000. The linear or two-dimensional ion-trap machines, such as the LTQ LC-MS/MS[®] from Thermo Finnigan in San Jose, California, are also configured to offer greater ion storage capacity.

The new 4000 Q TRAP LC-MS/MS system from Applied Biosystems in Foster City, California, and MDS Sciex in Concord, Canada, is a triple-quadrupole mass spectrometer where the last quadrupole can be operated as a conventional quadrupole mass filter or as a linear ion trap — an arrangement that allows decoupling of precursor ion isolation and fragmentation from the ion trap itself. The company says that this provides more sensitive tandem mass-spectrometry (MS/MS) and allows the identification and measurement of low levels of post-translational modification in a single liquid chromatography–MS/MS run. "This is one of the few instruments where you'd actually be able to do the protein identification work, protein characterization and absolute quantitation on the same instrument,"

claims Ron Bonner, director of applied research at MDS Sciex.

At the top end of the market, Bruker Daltonics in Billerica, Massachusetts, and Thermo Finnigan have hybrid Fourier transform MS (FTMS) instruments. Bruker's APEX-Q analyses ions in the ion cyclotron resonance cell under high vacuum in a high magnetic field. Frank Laukien, president and chief executive of Bruker Daltonics, says that the hybrid 'front-end' for FTMS, high-field magnets (9.4 or 12 tesla) and COMPASS software are suitable for 'shotgun' proteomics as well as a 'top-down' approach that does not depend on enzymatic digestion as a precondition for protein-sequence analysis. The Finnigan LTQ-FT consists of a linear ion trap, a high-transmission ion-guidance system, and an ion cyclotron resonance analyser based on a 7-T superconducting magnet.

But the ultra-high resolution and greater mass accuracy of FTMS comes at a price. The APEX-Q (9 T model) starts at about \$1.25 million and rises to \$2 million for the top-of-the-range 12-T model. And FTMS instruments have a reputation for being tricky to operate, although Laukien says that "anybody who can run an ion trap or a Q-TOF can also run a FTMS". Whether the enhanced analytical capability of these high-end instruments will offset their cost remains to be seen. John Yates of the Scripps Research Institute in La Jolla, California, believes it might. "I think these are going to have a huge impact if they wind up being user-friendly," he says.

D.G.

(see R. Aebersold and M. Mann *Nature* **422**, 198–207; 2003). “If you want to just do protein identification in fairly normal organisms, such as yeast, humans and mice, where the genomes are known, then that’s pretty straightforward,” says Ron Bonner, director of applied research at mass spectrometer developers MDS Sciex in Concord, Canada.

Caught in a trap

Mass spectrometers consist of an ion source, a mass analyser and a detector. Of the four main types of mass analyser currently in use (ion trap, time-of-flight, quadrupole, and Fourier transform ion cyclotron), the workhorse has been the standard three-dimensional (3D) ion trap. “They’re relatively cheap, they’re extremely robust and they have good performance, but the mass accuracy and mass resolution is really not all that great,” says Ruedi Aebersold of the Institute for Systems Biology in Seattle, Washington. “Sensitivity-wise, they’re actually very good.”

The more recent linear, or two-dimensional, ion traps (see ‘Mass spectrometry goes mainstream’, page 581) have better trapping efficiency and increased ion capacity, which gives them greater resolution, sensitivity and mass accuracy. Electrospray ionization, which ionizes the analytes out of a solution and so can be coupled to liquid-based separation systems, is used in combination with a variety of mass analysers, particularly ion traps. Time-of-flight (TOF) analysers measure the mass of intact peptides with high accuracy and resolution and are often used for high-throughput protein identification



Keith Williams says that providing start-to-finish solutions is key.

by peptide mapping (peptide mass fingerprinting). In these machines, the ionization method is typically matrix-assisted laser desorption/ionization (MALDI).

The novel Sensitizer peptide-labelling reagent developed by Proteome Sciences of Cobham, UK, is aimed at improving the sensitivity of MALDI-TOF and SELDI (surface-enhanced laser desorption/ionization)-TOF MS systems, in particular, without requiring modifications to instruments or data-analysis software. Increasing the number of proteins identified by peptide mass fingerprinting should mean that fewer samples require full *de*

novo sequencing using the more time-consuming tandem MS (MS/MS).

The buzz this year is over Fourier transform MS (FTMS), with two hybrid machines specifically targeted at proteomics coming onto the market (see ‘Mass spectrometry goes mainstream’, page 581). FTMS was used to identify more than 60% of the predicted proteome of the bacterium *Deinococcus radiodurans*, the most complete coverage of any proteome yet. And although the analysis of post-translational modifications such as glycosylation and phosphorylation using MS is still something of a challenge, FTMS was able to identify almost every post-translational modification of a 29 kDa protein.

The well-established technique of two-dimensional gel electrophoresis (2DE) has been used most commonly for separating protein mixtures in preparation for MS. And this is likely to continue in the near future. To some extent, “it’s because old habits die hard”, says Aebersold. After staining and image analysis, selected spots from the gel are excised, digested and the peptides analysed by MS. Despite attempts at automation (see ‘Look, no hands’, below), 2DE is likely to remain fairly low-throughput. It also requires relatively large amounts of sample (an issue with clinical material) and in the past has failed to detect low-abundance proteins reliably. Pre-fractionation of the sample, or the use of more sensitive staining methods and larger-format gels, can enhance detection of low-abundance proteins.

The shortcomings of 2DE–MS have fuelled interest in eliminating the need for gels. Liquid

LOOK, NO HANDS

A recurring theme in proteomics is the demand for greater automation to increase efficiency and reproducibility. But to improve somebody’s process, “you first have to convince them they’ve got one”, says John Hobbs, strategic market manager for Beckman Coulter in Fullerton, California. Beckman Coulter has modified and repositioned some of its products under the name ProteomeLab. The products are very “solutions-orientated”, says Hobbs. With the PF 2D automated, two-dimensional protein-fractionation system, for example, the company supplies the chemistry, the hardware, the software and the methods, and “we basically tell the customer don’t mess with it”, he says. The PF 2D uses chromatofocusing followed by non-porous reverse-phase chromatography for high-resolution separation of proteins from complex mixtures. Fraction collection from the first dimension is controlled by an in-line pH monitor, with automatic injection of the fractions into the second dimension. Liquid fractions can be stored or the eluent connected directly to electrospray ionization–mass spectrometry. Beckman Coulter is working with mass spectrometer vendors to ensure compatibility with MALDI (matrix-assisted laser desorption/ionization) plate spotters.

Although non-gel-based methods coupled with mass spectrometry (MS) are gaining acceptance, the most commonly used approach is still a combination of two-dimensional gel electrophoresis (2DE) and MS. Two-dimensional gel electrophoresis is difficult to automate, but NextGen Sciences of Huntingdon, UK, will soon be shipping a fully automated 2DE

platform — a2DE. Immobilized pH gradient strips (up to 30 cm) are run in parallel in the first dimension and three large-format SDS-PAGE gels in the second, and all steps between sample injection and SDS-PAGE are automated. Future versions will allow the same degree of optimization in the first dimension as currently offered in the second dimension, says chief executive Kevin Auton. Then “you can really start to zoom in on certain isoelectric points, as well as molecular-weight ranges”, he says.

For about US\$2 million, Proteome Systems in North Ryde, Australia, sells a start-to-finish automated proteomics solution called ProteomIQ, launched last year — or you can buy the components separately. It comprises kits and instruments for sample preparation and protein separation, Shimadzu-Biotech’s Axima CFR Plus MALDI-TOF (time-of-flight) mass spectrometer and bioinformatics capabilities. The Xcise gel-processing robot, developed jointly with Shimadzu-Biotech of Kyoto, Japan, bundles the functions of gel imaging, precision spot cutting, protein digestion and peptide purification, and deposition onto a target plate into one instrument. Also developed with Shimadzu-Biotech, the novel Chemical Inkjet Printer (ChIP) uses piezoelectric printing technology to dispense minute amounts of reagents onto selected protein spots electroblotted from 2D gels onto membranes. After processing, the membrane can be placed directly into the mass spectrometer for peptide analysis. ChIP not only cuts down on liquid-handling steps but also offers the option of archiving samples for future analysis.

D.G.

chromatography (LC)–MS/MS is gaining in acceptance and is generally more amenable to automation. If quantification is required, the proteins or peptides can be labelled with stable isotopes to enable measurement of differential levels of protein expression. One option is the site-specific, covalent labelling of proteins with isotope-coded affinity tags such as the ICAT reagents from Applied Biosystems in Foster City, California.

Technologies for analysing biomolecular interactions, such as surface plasmon resonance (SPR), are also hitching themselves up to MS. BIAcore of Uppsala, Sweden, has integrated its BIAcore 3000 SPR-based instrument with MALDI–MS analysis so that proteins bound to the ligand on the sensor surface can be automatically recovered and deposited directly onto the MALDI plate for further analysis by MALDI–TOF or TOF/TOF–MS.

Good preparation

But no matter what the analytical method, good sample preparation is essential. Fractionation methods such as centrifugal ultrafiltration, immunoaffinity purification and free-flow electrophoresis (FFE) can be used to reduce the complexity of the sample and enrich rare proteins before processing by 2DE or LC–MS. Tecan in Männedorf, Switzerland, sells the ProTeam FFE workstation, a matrix-free FFE fractionation system for processing proteins, organelles, membrane fragments or whole cells. The sample flows continuously in a thin film of aqueous medium between two parallel plates. The application of a high

voltage generates an electric field perpendicular to the laminar flow. Charged species are deflected, allowing fractions to be collected on a fast, preparative and continuous basis.

The Gyrolab MALDI SP1 from Gyros in Uppsala, Sweden, is a microlaboratory in the form of a compact disc on which sample preparation procedures (sample clean-up and concentration) for peptide mapping or sequencing by MALDI–MS are miniaturized and integrated. Gyros also sells the MALDI IMAC for concentrating, purifying and crystallizing phosphorylated peptides directly onto MALDI target areas on the CD. Gyrolab workstations and kits are available for use with MALDI mass spectrometers from Shimadzu-Biotech in Kyoto, Japan, and Bruker Daltonics in Billerica, Massachusetts. MassPREP target plates from Waters-Micro-mass of Milford, Massachusetts, are designed for use with the company's robotic protein-handling system (MassPREP Station) and the Waters MALDI–TOF family of mass spectrometers. They provide fast, simple 'on-plate' preparation of protein digests and are designed to offer a tenfold increase in sensitivity over conventional target plates.

Several companies now sell automated equipment for spotting proteins onto MALDI targets. The Microlab Star MALDI Spotting Workstation from Hamilton Bonaduz, Switzerland, for example, has a volume range of 0.5–1,000 microlitres and can spot a 384-well target in 1 or 2 hours, depending on the number of channels fitted. The Ettan Spot Handling Workstation 2.1

from Amersham Biosciences in Piscataway, New Jersey, automates the process from spot picking from 2D gels through to digesting, drying and spotting the resultant peptides onto the target plates. A sample loader that will automatically place the target plates into the mass spectrometer is in the works.

With many of the component techniques now automated, the next trick is to bring them together. Start-to-finish automated proteomics systems are beginning to appear. Last year, Proteome Systems launched ProteomIQ (see 'When the chips are down', below). Others hope to do likewise. Amersham Biosciences has joined forces with Thermo Electron in Waltham, Massachusetts, and Waters-Micro-mass has partnered with Bio-Rad of Hercules, California, to develop ProteomeWorks.

Alternative angles

A quite different approach to probing protein activity and function is the protein microarray, of which there are two basic types. The analytical microarray contains an ordered array of protein-specific ligands, typically antibodies, spotted onto a derivatized solid surface. They can be used to monitor differential protein expression, protein profiling and clinical diagnostics. But progress here is constrained by a lack of comprehensive sets of high-specificity, high-affinity antibodies.

UK-based Cambridge Antibody Technology is tackling this problem by using phage display to create large libraries of antibodies

WHEN THE CHIPS ARE DOWN

Off-the-shelf protein arrays are just beginning to trickle onto the market. But most have been low-density antibody arrays for profiling cytokines, secreted proteins that act as intercellular chemical messengers. Panomics of Redwood City, California, produces TranSignal Human Cytokine Antibody Array 2.0 for profiling the expression of 21 well-studied human cytokines. The array uses standard chemiluminescence detection and detects soluble cytokines at concentrations in the picogram per millilitre range. Zymyx of Hayward, California, offers a human cytokine biochip for profiling 30 cytokines, which is sold for use with its Protein Profiling Biochip System, which includes an automated workstation, biochip reader and data-analysis software.

BD Biosciences Clontech in Palo Alto, California, sells an off-the-shelf antibody array for detecting a variety of cytosolic and membrane-bound proteins relevant to signal transduction, cell-cycle regulation, gene transcription and apoptosis. Its Antibody Microarray 500 contains over 500 antibodies printed on a standard-size (75 × 25 × 1 mm) glass slide and uses fluorescence-based detection. It is designed for qualitative rather than quantitative analysis but can detect differences in protein abundance between two samples with detection limits of 20 pg ml⁻¹ for each protein target.

On the protein-function microarray front, Protometrix of Branford, Connecticut, will later this year roll out its first product, the Yeast ProtoArray, which stems from the pioneering protein microarray work of Michael



The Yeast ProtoArray from Protometrix contains some 5,000 proteins.

©PROTOMETRIX INC., USED WITH PERMISSION

Snyder's group at Yale University. This 'proteome' microarray will contain some 5,000 proteins in the *Saccharomyces cerevisiae* (yeast) proteome, including over 1,500 with significant homology to human proteins.

"The biggest bottleneck for everybody is still content," says Barry Schweitzer, Protometrix's senior director for technology. The company has now industrialized the cloning, expression and protein-purification process. "It's just a factory running 24/7," he says. "Everything that we clone we also have to sequence." The Yeast ProtoArray is targeted more towards the research market but Protometrix expects to follow this up with products of direct relevance to drug discovery.

D.G.

with the desired specifications. The company is also developing a cell-free system for producing antibodies using 'ribosome display' technology, which may make it possible to build larger libraries than with phage display.

Other companies are exploring capture agents that can be generated more efficiently than antibodies and are more stable to heat or changes in pH. These include protein scaffolds, such as Trinetin binding proteins from Phyllos in Lexington, Massachusetts, and Affibody affinity ligands from Affibody in Stockholm; RNA or DNA aptamers — oligonucleotide sequences — from SomaLogic of Boulder, Colorado; ribozymes (Riboreporters) from Archemix in Cambridge, Massachusetts; and partial-molecule polymeric imprints (ProteinPrint films) from Aspira Biosystems in South San Francisco, California. Commercial protein microarrays are starting to appear on the market, but so far most are relatively low-density antibody arrays for profiling cytokines.

Functional fit

A second type of protein microarray is the functional protein microarray, with potential for the high-throughput analysis of whole proteomes or other large collections of proteins or protein domains. Gavin MacBeath, assistant professor in the department of chemistry and chemical biology at Harvard University, is using such microarrays to help understand how the cell regulates complex processes such as apoptosis, growth-factor signalling, intercellular communication and protein trafficking.

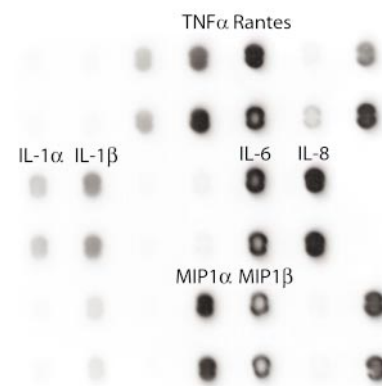
But it is slow going, he says. Because there are no large, cheap, defined sets of cloned genes available, he has to clone the genes and then express and purify the proteins required for the arrays. MacBeath arrays collections of proteins or protein domains onto a variety of chemically derivatized glass slides such as those from TeleChem International in Sunnyvale, California, typically using a home-built arrayer or a machine from Affymetrix in Santa Clara, California. Standard fluorescent dyes are used as labels and slides are scanned using either an ArrayWoRx scanner from Applied Precision in Issaquah, Washington, or a GenePix 4000B microarray scanner from Axon Instruments in Foster City, California.

Most papers to date on protein arrays have been 'proof-of-concept', he says. "We're at the point now where people need to start applying this technology to real questions and use it to learn something about biology."

There are now tools on the market to make array technology more accessible and higher throughput. PerkinElmer Life Sciences in Boston, Massachusetts, for example, sells the ProteinArray Workstation, which automates the processing of multiple protein microarrays. Developed in conjunction with NextGen Sciences, the system can process up



Gavin MacBeath uses high-throughput microarrays to probe protein function.



Panomics' cytokine microarray.

to 48 protein microarrays in 1–2 hours at maximum capacity.

There has been considerable interest in recent years in applying proteomics to clinical diagnostics and predictive medicine. The goal is to identify disease markers, or biomarkers, that can be used to extract diagnostic (or even prognostic) information from body fluids such as serum, saliva or urine.

The SELDI-based ProteinChip System from Ciphergen Biosystems of Fremont, California, for example, can distinguish between patients with ovarian cancer and women without the disease. The approach seems promising as a means of early detection, but there are questions about the robustness and reproducibility of the serum proteomic patterns generated, which should be

answered after larger-scale clinical studies.

Serum is applied directly to the surface of a ProteinChip Array and proteins are selectively retained on a variety of chromatographic surfaces. After incubation, unbound proteins are washed off, an energy-absorbing matrix is applied, and the array is analysed by TOF-MS. Most recently, Ciphergen has changed the way it makes its arrays. "It's our goal to become a 'matrix-free' company," says Scott Weinberger, who runs Ciphergen's proteomics research group. Last month it launched a new line of Surface Enhanced Neat Desorption (SEND) ProteinChip Arrays, in which the matrix is incorporated in the surface chemistry of the array. This not only enhances throughput, but also improves sensitivity and sequence coverage. SEND ID, the company's first SEND product, is targeted at peptide mass fingerprinting applications.

To unravel the complexity and capture the dynamic nature of the human proteome, researchers will need to have all the tools in the proteomics tool-box at their disposal as no one technology will suffice. Making sense of the data will require sophisticated IT solutions to analyse, manage and exchange data and here proteomics lags some way behind genomics and gene-expression studies. But, ultimately, it will require the integration of proteomics data with data from gene expression studies and functional assays, as well as with structural information on proteins and protein complexes provided by X-ray crystallography and nuclear magnetic resonance spectroscopy. ■

Diane Gershon is assistant editor, *Nature Medicine* Technical Reports.

Company	Products/activity	Location	URL
Protein preparation and mass spectrometry			
Advion BioSciences	Intellisense ESI chip for MS; contract liquid chromatography	Ithaca, New York	www.advion.com
Agilent	Mass spectrometers; 1100 series systems for HPLC; capillary electrophoresis systems; workstation for RNA, DNA or protein analysis	Palo Alto, California	www.agilent.com
Amersham Biosciences	Instruments, equipment and products for proteomics and cellular assays; Ettan range of equipment for sample preparation and mass spectrometry	Uppsala, Sweden	www5.amershambiosciences.com
Applied Biosystems	Protein sequencers, mass spectrometers, chromatography systems, peptide synthesizers; ICAT stable-isotope labeling reagents	Foster City, California	home.appliedbiosystems.com
ARRM	Robotic spot cutter and plate replicator	Adelaide, Australia	www.arrm.com
Beckman Coulter	Automated tools for molecular biology, biochemistry, genomics and proteomics; ProteomeLab	Fullerton, California	www.beckmancoulter.com
Bio-Rad	Products, instruments and software for life-sciences research; ProteomeWorks integrated system; WorksBase bioinformatics software	Hercules, California	www.bio-rad.com
Bruker Daltonics	Instruments, software and consumables for mass spectrometry for the life sciences; Proteiner proteomics software suite	Billerica, Massachusetts	www.bdal.com
Gyros	CD microlaboratory for protein preparation for mass spectrometry	Uppsala, Sweden	www.gyros.com
Hamilton Bonaduz	Microlab workstations for liquid-handling, MALDI spotting and protein crystallization	Bonaduz, Switzerland	hamiltoncomp.com
Intrinsic Bioprobes	Proteomics products and analysis based on proprietary BIA/MS technique	Tempe, Arizona	www.intrinsicbio.com
KBiosystems	Automated equipment for proteomics	Basildon, UK	www.kbiosystems.com
MDS Sciex	Mass spectrometer developers and producers	Ontario, Canada	www.sciex.com
Micromass-Waters	Micromass instruments for mass spectrometry; HPLC systems; reagents and consumables	Milford, Massachusetts	www.waters.com
Millipore	Automated equipment for life sciences research; Montage In-Gel DigestZP Kit and ZipPlate Micro-SPE plates for protein sample preparation from gel for MS	Bedford, Massachusetts	www.millipore.com
NextGen Sciences	a2DE automated 2D-gel electrophoresis workstation	Huntingdon, UK	www.nextgensciences.com
Proteome Systems	Proteomics technology; ProteomIQ integrated system for proteomics	North Ryde, Australia	www.proteomesystems.com
Qiagen	BioRobot multifunctional laboratory workstation and equipment for automated nucleic acid, protein and cell purification; reagents, kits and instruments for proteomics	Venlo, the Netherlands	www.qiagen.com
RoboDesign	Automated equipment for biotechnology, protein structure analysis, genomics and proteomics	Carlsbad, California	www.robodesign.com
Shimadzu-Biotech	AXIMA mass spectrometers; XCISE integrated sample gel excision processor; Chemical Inkjet Printer	Kyoto, Japan	www.shimadzu-biotech.net
Thermo Finnigan	Instruments for chromatography and mass spectrometry; ProteomeX workstation for automated LC-MS protein analysis	San Jose, California	www.thermo.com
Protein arrays			
Advantix	ArrayBooster nanopump system for protein chip and microarray incubation	Munich, Germany	www.advantix.de
Affibody	High-throughput protein identification and characterization by proprietary binding proteins	Stockholm, Sweden	www.affibody.com
Affymetrix	Spotters and arrays	Santa Clara, California	www.affymetrix.com
BioGenex	Automated DNA and protein chip processing and scanning	Ramon, California	www.biogenex.com
HTS Biosystems	FLEX CHIP array platform for detection of protein binding using surface plasmon resonance	Hopkinton, Massachusetts	www.htsbiosystems.com
Molecular Probes	Fluorescent probes	Eugene, Oregon	www.probes.com
Molecular Staging	Proprietary RCAT chip-based proteomics platform; whole-genome amplification kits	New Haven, Connecticut	www.molecularstaging.com
NanoPlex Technologies	Nanobarcode metallic nanoparticle detection tags for proteomics and other applications	Mountain View, California	www.nanoplextech.com
Panomics	TransSignal Human Cytokine Antibody Array for profiling expression of 18 human cytokines	Redwood City, California	www.panomics.com
PerkinElmer Life Sciences	Automated systems for liquid handling and sample preparation; HydroGel BioChip; ProteinArray workstation for protein arrays	Boston, Massachusetts	lifesciences.perkinelmer.com
Phylos	Trinectin binding proteins for protein capture	Lexington, Massachusetts	www.phylos.com
Protomatrix	Yeast protein microarray for identification and interaction studies	Branford, Connecticut	www.protomatrix.com
Scienion	sciCAPTURE customized protein array; sciLISA antibody array	Berlin, Germany	www.scienion.de
SomaLogic	Photoaptamer arrays for proteomics	Boulder, Colorado	www.somallogic.com
TeleChem/ArrayIt International	Protein Edition Spot Bot for making protein arrays	Sunnyvale, California	www.arrayit.com
Zeptosens	ZeptoMARK protein chip platform using planar optical waveguide technology	Witterswil, Switzerland	www.zeptosens.com
Zyomyx	Human cytokine profiling biochip system	Hayward, California	www.zyomyx.com
Protein crystallization and structure determination			
BSI Proteomics	Proprietary system for automated protein crystallization for crystallography	Gaithersburg, Maryland	www.bsipteomics.com
DataCentric Automation	Automated systems for protein crystallography	Nashville, Tennessee	www.titanceg.com
Fluidigm	Microfluidics chip for protein crystallization	South San Francisco, California	www.fluidigm.com
Gilson	Automated liquid-handling, pipetting and protein crystallization equipment	Middleton, Wisconsin	www.gilson.com

Company	Products/activity	Location	URL
Greiner Bio-One	Consumables for high-throughput research and diagnostics; microplates; Crystal Star microplate platforms for protein crystallization	Frickenhausen, Germany	www.greinerbioone.com
Upstate	KinaseProfiler specificity testing, IC50Profiler Express and services for crystallography assay development, bulk protein production and purification	Charlottesville, Virginia	www.upstate.com
Inpharmatica	Biopendium and CelerEdition Biopendium proteome annotation resources; PharmaCarta computational genomics system	London, UK	www.inpharmatica.com
General			
20/20 GeneSystems	P-FILM Layered Gene Scanning membrane system for protein or nucleic acid capture	Rockville, Maryland	www.2020gene.com
Applied Precision	arrayWoRxe biochip reader	Issaquah, Washington	www.appliedprecision.com
Apogent Technologies	Labware and equipment for the life sciences	Portsmouth, New Hampshire	www.apogent.com
Antibodies by Design/ MorphoSys	Recombinant antibody library for custom generation of monoclonal antibodies	Martinsried, Germany	www.antibodyservices.com ●
Axon Instruments	GenePix 4000B microarray scanner and GenePix Pro microarray analysis software	Foster City, California	www.axon.com
Becton Dickinson/ BD Biosciences	Culture media, radioassay kits, FACS range of flow cytometers; monoclonal antibodies, antibody arrays for proteomics; reagents and biochemicals for molecular biology	Franklin Lakes, New Jersey	www.bd.com
Biacore	Analysis and measurement of biomolecular interactions using surface plasmon resonance	Uppsala, Sweden	www.biacore.com
BioVisioN	Peptidomics: technologies for discovery and identification of peptides and small proteins	Hannover, Germany	www.biovision-discovery.de
BMG Labtechnologies	Microplate and array readers	Offenburg, Germany	www.bmg-labtechnologies.com ●
brainprot	Proteomics analysis	Vienna, Austria	gert.lubec@akh-wien.ac.at ●
Brinkmann	Laboratory instrument suppliers; software; consumables	Westbury, New York	www.brinkmann.com
Chromagen	Fluorescence based products, kits and services for gene expression and protein assay	San Diego, California	www.chromagen.com
Complexio	Proteome analysis of biological membranes, identification and characterization of membrane microdomains, and protein interactions	March, Germany	www.complexio-gmbh.com ●
Cytomx	1- and 2D gel analysis, mass spectrometry and amino-terminal sequencing for proteins	Cambridge, UK	www.cytomx.com
Dynal	Dynabeads for immunoprecipitation and phage-display panning	Oslo, Norway	www.dynalbiotech.com
Eppendorf	Liquid-handling equipment, centrifuges	Hamburg, Germany	www.eppendorf.com
Eurogentec	Large-format 2DE; protein identification by MALDI, ESI/MS/MS, Edman sequencing	Seraing, Belgium	www.eurogentec.com
GeneProt	Industrial-scale separation, identification and characterization of human proteins	Geneva, Switzerland	www.geneprot.com
Genetix	Picking, arraying, gridding and replicating robots for genomics and proteomics	New Milton, UK	www.genetix.co.uk ●
Genomic Solutions	Investigator automated system for proteomics; software, consumables and services	Ann Arbor, Michigan	www.genomicsolutions.com
Hybrigenics	<i>In silico</i> prevalidation, and experimental validation of drug targets	Paris, France	www.hybrigenics.com
Invitrogen	Kits and reagents for proteomics, molecular and cell biology	Carlsbad, California	www.invitrogen.com
Jerini Peptide Technology	High-density peptide microarrays; kinase identification and substrate profiling	Berlin, Germany	www.jerini.com ●
Novagen	Reagents for genomics and proteomics	Madison, Wisconsin	www.novagen.com
Peptide Specialty Laboratories	Custom peptides	Heidelberg, Germany	www.peptide.de ●
Procognia	Glycoanalysis of glycoproteins	Maidenhead, UK	www.procognia.com
Prodigene	Protein production in plants; vaccine development	College Station, Texas	www.prodigene.com
Promega	Vectors, reagents and kits for genomics, proteomics and cellular analysis	Madison, Wisconsin	www.promega.com
Protagen	Protein analysis services	Dortmund, Germany	www.protagen.com
Proteome Sciences	Drug target and disease biomarker discovery; ProteoSHOP toolkit for high-output proteomics	Cobham, UK	www.proteome.co.uk
Roche Diagnostics	Reagents and kits for molecular biology, functional genomics and proteomics research	Lewes, UK	www.roche-applied-science.com
Schleicher & Schuell	Slides, kits, equipment and reagents for protein microarrays	Keene, New Hampshire	www.schleicher-schuell.com/bioscience
Stratagene	Tools and reagents for proteomics	La Jolla, California	www.stratagene.com
Tecan	Automated solutions for proteomics: ProTeam free-flow electrophoresis for fractionation, automated 2D-PAGE spot picking, in-gel digester and interface to mass spectrometer	Männedorf, Switzerland	www.tecan.com ●
Trenzyme	Genomics and proteomics services: cloning, mutagenesis, library construction, gene expression, protein purification	Konstanz, Germany	www.trenzyme.com ●
Xzillion	High-throughput protein profiling using protein sequence tag technology	Frankfurt, Germany	www.xzillion.com
Databases and software			
Accelrys	Discovery Studio products for database mining, genomics and proteomics	San Diego, California	www.accelrys.com
Applied Maths	Gel fingerprint analysis and bioinformatics software; bioinformatics contract services	Kortrijk, Belgium	www.applied-maths.com ●
Array Genetics	Online protein information database and tools for genomics, proteomics, microarray analysis	Newtown, Connecticut	www.arraygenetics.com
Compugen	GenCarta annotated human genomic, transcriptome and proteome database	Tel-Aviv, Israel	www.cgen.com
DECODON	Software for 2D-gel analysis and information storage	Greifswald, Germany	www.decodon.de
GeneBio	Commercial access to SWISS-PROT and other Swiss Institute of Bioinformatics databases	Geneva, Switzerland	www.genebio.com
Incyte Genomics	Proteome BioKnowledge Library species-specific protein information databases; LifeSeq Foundation database; bioinformatics software	Palo Alto, California	www.incyte.com
Informax	Software for genomics and proteomics	Bethesda, Maryland	www.informaxinc.com
LION Bioscience	Software for genomics and proteomics	Heidelberg, Germany	www.lionbioscience.com
Matrix Science	MASCOT search engine for mass spectrometric data searches against sequence databases	London, UK	www.matrixscience.com
MegaMetrics	DIOGENES data-mining program for analysis of microarray, proteomics and SNP databases	Wyndmoor, Pennsylvania	www.megametrics.com
Nonlinear Dynamics	Image-analysis and database software for gel electrophoresis and high-throughput arrays	Newcastle-upon-Tyne, UK	www.nonlinear.com
Proteomic Solutions	Melanie software for 2D gel analysis; distributor for Genomic Solutions; proteomics services	St Marcel, France	www.proteomicsolutions.fr
PubGene	PubGene databases and analysis software for genomics and proteomics	Oslo, Norway	www.pubgene.com
Structural Bioinformatics	Protein and protein-structure databases; computational proteomics for drug discovery	San Diego, California	www.strubix.com

● see advertisement

Contacts

Publisher: Ben Crowe
Editor: Paul Smaglik
Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Global Head of Naturejobs:
Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)
Andy Douglas (4975)
Frank Phelan (4944)

Netherlands/ Italy/ Iberia/ Belgium:

Evelina Rubio Hakansson (4973)
Scandinavia: Silje Opstrup (4994)

France/ Switzerland:
Amelie Pequignot (4974)

Natureevents:
Paul Constant (4954)

Production Manager:

Billie Franklin
To send materials use London
address above.
Tel +44 (0) 20 7843 4814
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

International

Advertising Coordinator:
Hind Berrada (4935)

Naturejobs web development:
Tom Hancock

Naturejobs online production:
Ben Lund

European Satellite Office

Germany/ Austria:
Patrick Phelan, Odo Wulffen
Tel + 49 89 54 90 57 11/2
Fax + 49 89 54 90 57 20
e-mail: p.phelan@nature.com
o.wulffen@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707
Tel +1 800 989 7718
Fax +1 800 989 7103
e-mail: naturejobs@nature.com

US Sales Manager: Peter Bless

US Advertising Coordinator:
Linda Adam

Japan Head Office, Tokyo

MG Ichigaya Building (5F),
19-1 Haraikatamachi,
Shinjuku-ku,
Tokyo 162-0841
Tel +81 3 3267 8751
Fax +81 3 3267 8746

Asia-Pacific Sales Director:
Rinoko Asami
e-mail: rasami@nature.com

naturejobs

Redressing the balance

The figures are all too familiar. In the United States, African Americans and Hispanics together make up only 4.5% of scientists with a PhD in the biological sciences, and 6% of practising physicians. These numbers have been historically resistant to change. But summer research programmes that put minority students into labs with mentors may help to turn the tide.

One such programme pairs up researchers from Columbia University College of Physicians and Surgeons with minority undergraduates from Hunter College in New York City. The 16 students are spending ten weeks in the lab with their mentors. Apart from doing research, they will be trained in research design and analysis, and will be exposed to scientific literature through journal clubs. They will also meet minority members who have been successful in science and business — Cecil Pickett, president of Schering-Plough Research Institute in Kenilworth, New Jersey, will this week visit the students. Pickett has been named as one of *Fortune's* 'most powerful black executives'.

Although the mentoring programme is small and only began last year, it is already showing signs of making an impact. Several of last summer's class have been accepted at medical schools and on graduate science programmes. The remainder are continuing their undergraduate education at Hunter.

To change the statistics of minority representation, more such programmes are needed. Ideally, they would include high-school students, or even younger potential scientists. They would encompass all disciplines and bring in scientists from a variety of professional backgrounds. Exposing students to a range of scientists who are excited about their work and able to communicate the career possibilities is the best way to interest students in science, regardless of either group's ethnicity.

Paul Smaglik
Naturejobs editor



Contents

CAREERS AND RECRUITMENT

Sharpening the skills of
medicinal chemists p594

The changing needs
of drug discovery p597

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

ANNOUNCEMENTS

EVENTS

Life beyond synthesis

Medicinal chemists have skills that are eagerly snapped up by industry, and an increase in training options is making it easier to gain the necessary experience. Eugene Russo reports.

Fresh perspective: medicinal chemists need to be able to think both at the molecular level and at a more global, descriptive level — a key part of their training.



ASTRAZENECA

Today's medicinal chemist is the pharmaceutical world's equivalent of a chef trying to create the perfect soufflé. He or she must come up with the best recipe for the available ingredients, and then refine it through trial, error and intuition.

Traditionally, medicinal chemists have been trained as synthetic organic chemists and their job has been to synthesize the molecules that might eventually become new drugs, but their job descriptions — and skill sets — are beginning to change (see 'Changing definitions', opposite). There are also new sources of employment. Once primarily the domain of pharmaceutical companies, medicinal chemistry has become more common at larger biotech companies and some academic research departments (see page 597).

To adapt to these changes, several training options are available for scientists learning to become medicinal chemists. Additional training for young chemists and other medicinal chemist hopefuls is increasingly necessary as, more and more, medicinal chemistry draws on wider sources of information, such as pharmacology and structural biology. As well as the traditional on-the-job training, there is an

increasing number of specialist short courses, ranging from a week to several months long. In addition, there are some institutes dedicated specifically to medicinal-chemistry research and teaching, such as the Drug Research Academy at the Danish University of Pharmaceutical Sciences in Copenhagen.

LEARNING CURVE

According to a survey in the mid-1990s carried out by the International Union of Pure and Applied Chemistry, most big drug companies tend to hire organic and synthetic chemists and then train them to be medicinal chemists on the job. Companies want experts in compound synthesis — knowledge of biology is important, but secondary.

The study's author, Lester Mitscher, a medicinal chemist at the University of Kansas, says that this has much to do with the observation that chemists can learn biology but biologists have more difficulty learning chemistry. "Chemists automatically think in molecular terms and biologists are much more likely to think in more global or descriptive terms," says Mitscher. "And it's easier to make the transition

from molecular to global than to go the other way.”

Availability has also influenced hiring strategies; there are many more organic- and synthetic-chemistry departments at universities than there are medicinal-chemistry departments. At university, students learn how to construct molecules, but they don't get much training in which molecules are worth making, says Mitscher.

Medicinal chemist William Greenlee, a vice-president at the Schering-Plough Research Institute in Kenilworth, New Jersey, cites simple tradition as feeding the tendency to recruit synthetic chemists — medicinal chemists tend to perpetuate the hiring practices that were the basis for their own careers. “I've always considered it a sort of trial by fire,” he explains. “Typically we hire people who have only synthetic organic experience, and then they have a fairly steep learning curve when they join to learn about basic medicinal chemistry.”

A BROADER VIEW

But Greenlee suggests that it may be time for companies to expand their horizons, to look to candidates with a background in medicinal chemistry or some training beyond synthetic organic chemistry. With the current focus on high-throughput technologies of combinatorial chemistry and parallel synthesis, and with drug leads now rarely coming from natural products, he points out that the types of molecule made are often not as complex as they once were. So the exclusive emphasis on the experienced chemist with training in the total synthesis of natural products may be somewhat misplaced.

University chemistry curricula are increasingly including courses in medicinal chemistry or, at the very least, research projects with a medicinal-chemistry focus. “The difficulty with that is that the professors themselves have to learn enough biology to cope. Some do that exceptionally well, some do that very poorly,” says Mitscher.

Amendments to the curriculum are not necessarily

going to transform an organic chemist into a productive medicinal chemist. In the past 20 years, short courses — primers on the biology, physiology and pharmacology necessary for drug design and synthesis — have become crucial to training young chemists hired by industry. Britain's Royal Society of Chemistry held the first such course in 1979. In 1987, Drew University in Madison, New Jersey, located in the heart of a major centre of biotechnology and pharmaceuticals, started its one-week course in medicinal chemistry.



William Greenlee:
medicinal chemistry can
be a steep learning curve.

Short courses on this model have since been established in many countries, including Switzerland, India, Italy and the Netherlands. A course affiliated with the University of California, San Diego, started in 2001, and the university also runs a fuller 18-week evening course for people working in local industry.

These short courses are meant to fill in the knowledge gaps of aspiring medicinal chemists and accelerate their training. The course at Drew, for example, features 18 one-hour lectures, three two-hour seminars, and four half-hour discussion sessions. Study topics range from drug

design to molecular modelling.

The course also includes case studies illustrating how compounds now close to approval by the US Food and Drug Administration were developed. These are presented by the researchers themselves, who discuss issues of biological activity and positive and negative clinical data to help participants learn valuable rules of thumb for drug design.

John Primeau, vice-president of infection chemistry at AstraZeneca R&D in Boston, has seen several new employees benefit from the course. “It jump-starts them,” he says. Although synthetic organic chemists are good at “putting molecules together,” says Primeau, they're not necessarily good at “putting a medicinal-chemistry plan together”, which involves making the right compounds to get the data necessary to proceed to the next step. As an alternative to these

Changing definitions

In the 1950s, professor emeritus Alfred Burger of the University of Virginia defined the principles of medicinal chemistry when he wrote its first-ever textbook, *Burger's Medicinal Chemistry and Drug Discovery*, now in its sixth edition. Burger also became the first editor of the *Journal of Medicinal Chemistry* at its foundation in 1958.

The field of medicinal chemistry today may have the same fundamental mission — designing molecules to meet specific biological objectives — but it has evolved, along with the new technologies, to become more interdisciplinary.

Salvador Moncada, director of the Wolfson Institute for Biomedical Research at University College London, agrees with the classic definition of a medicinal chemist as someone who uses organic synthetic chemistry to make compounds to probe biological mechanisms. But he warns that the definition shouldn't be restrictive, as other specialists, such as those proficient in microarrays or mass spectrometry, can have a huge role in molecule design.

Philip Portoghese, professor of medicinal chemistry at the University of Minnesota and current editor-in-chief of the *Journal of Medicinal Chemistry*,

says that the power of medicinal chemistry comes not from any one field, but from the interaction between several, including biochemistry and pharmacology. “It's really the relationship between molecular structure and biological activity,” he says.

For John Primeau, head of infection chemistry at AstraZeneca R&D in Boston, a medicinal chemist's greatest skill is in digesting and compiling data. He describes a medicinal chemist as someone who takes biological data and physical-property data to build a molecule that is then tested and generates more data, upon which the next version can be based. **E.R.**



ASTRAZENECA

Internal investment: AstraZeneca brings lecturers in-house to teach recruits the necessary skills to be medicinal chemists.

short courses, some companies such as AstraZeneca also bring lecturers in-house.

The course at Drew typically gets about 300 applicants for 150 places. The founder of the course, retired medicinal chemist William Houlihan, says that the programme intentionally accepts the least experienced chemists as they have the most to learn. Drew also tries to accept at least one scientist from each company that applies, although Houlihan says this is becoming more and more difficult with the increasing number of applications from biotech companies.

The current popularity of courses may indicate that an increasingly competitive job market has made additional training more attractive. "When people see their neighbours being laid off, they are much more concerned about broadening their knowledge and doing ongoing training," says Michael Griffith, who coordinates two medicinal-chemistry courses affiliated to the University of California, San Diego.

EXTRA SKILLS

Despite the success of short courses, they cannot offer the breadth of training that a course lasting several years will do. In some countries, new models may be needed. Copenhagen's Drug Research Academy, established in 2002, offers an interdisciplinary graduate education drawing on resources from academia and industry.

Medicinal chemist Povl Krogsgaard-Larsen of the Danish University of Pharmaceutical Sciences initiated the project after representatives of Denmark's pharmaceutical industry told him that they needed scientists trained in the wider aspects of drug discovery, including toxicology



John Primeau sees data assimilation as key.

and pharmacology. The academy has joint funding until 2007 from the university, the Danish government and industry.

Once at full capacity, the academy will fund 36 PhDs and 12 postdocs (six from Denmark, six from other countries); roughly half of these will focus on medicinal chemistry. The programme has prompted interest from elsewhere in Europe about setting up similar training programmes in medicinal chemistry.

Other collaborations between industry and academia are already attempting to bridge the training and research gaps. One example is the Wolfson Institute for Biomedical Research at University College London (UCL) founded in 1995. "When we came here to UCL, the idea of having organic synthetic chemistry and medicinal chemistry in a biology department wasn't very fashionable," says Salvador Moncada, the Wolfson's director. "Now everybody talks about it."

The institute has a built-in mechanism for cooperation with industry. At the beginning, Moncada and the institute's other founders set up a holding company that handles all the intellectual property generated by the institute's basic research; UCL owns half of the company, a trust controls the other half. Four companies have been spun off since the institute's foundation, including NCE Discovery, based at the institute, which offers medicinal-chemistry services.

Collaborations between research and academia, whether for training or research, or both, are likely to grow. But Mitscher warns that balancing academic and industry interests is crucial for finding the right mix of training options.

Eugene Russo is a freelance science writer in Takoma Park, Maryland.



Screening process: the links between chemistry, biology and medicine are getting stronger.

ASTRAZENECA

F. ERIKSSON

Success in an uneven market

Medicinal chemists are in demand more than ever before — but the vagaries of the drug-discovery market and a sluggish economy have dampened prospects temporarily, says Eugene Russo.

The tables have turned for jobs in medicinal chemistry. Traditionally, such skills were in demand at pharmaceutical companies, where medicinal chemists would be expected to come up with new molecules. Biotechnology companies, by contrast, preferred molecular biologists who could search for protein-based therapeutics. But, helped by several economic factors, these distinctions are now becoming blurred.

Demand for medicinal chemists in the drug industry has slowed somewhat, and signals are mixed in the biotech sector. Mergers in both sectors, as well as a shaky stock market, have dampened opportunities all round.

Although medicinal chemists are still in demand at big drug companies, the job market has become less buoyant compared with a year or two ago. "It's still a pretty good market," says Philip Portoghese, editor-in-chief of the *Journal of Medicinal Chemistry* and a medicinal chemist at the University of Minnesota. "But it's not what it was."

Joel Huff, a vice-president of medicinal chemistry based at Merck's facility in West Point, Pennsylvania, agrees. Merck has done minimal hiring worldwide in the past year, he says, partly as a result of the sagging economy, but mostly because of an increasing focus on the preclinical stages of drug development, which does not typically involve medicinal chemists.

But it is not all doom and gloom. John Primeau, a vice-president at AstraZeneca R&D in Boston, says that his company is expecting "steady growth" worldwide in terms of new jobs.

And over the past few years, chemistry has moved to the heart of many companies in the biotech sector. "There definitely has been a shift," says Richard Scheller, senior vice-president of research at Genentech in South San Francisco, California. The early biotech firms focused on developing protein therapeutics using recombinant DNA and gene-expression technologies. Now, companies are more likely to be founded on

small-molecule drug development than protein therapeutics — which has meant an increased demand for medicinal chemists who can design small molecules.

"These days more than 90% of biotech companies are orienting themselves towards the drug arena," says Povl Krogsgaard-Larsen, a medicinal chemist at the Danish University of Pharmaceutical Sciences in Copenhagen. "This makes medicinal chemistry a key discipline."

FREEDOM

Drug-seeking biotech firms offer more flexible, less bureaucratic environments that can foster the sort of risk-taking that is unlikely to be found at large pharmaceutical companies. "In the past few years, drug discovery has been too dominated by screening technologies and robotic technologies," says Krogsgaard-Larsen.

Safi Bahcall, chief executive of Synta Pharmaceuticals in Lexington, Massachusetts, sees a major advantage in doing medicinal chemistry at a smaller company. "It's exciting seeing a drug get to patients," he says. "Do you want to be employee 99,950, or do you want to see a drug go into the clinic from start to finish?"

But few biotech companies have yet to attain profitability. "It's a game of high risk, high return," says Per Lindell, the vice-president of business development at Avalon Pharmaceuticals, a biotech firm based in Germantown, Maryland.

Many chemistry-based companies are currently in a holding pattern. Biovitrum, for instance, a 580-employee firm in Stockholm, hopes eventually to have the capacity to do its own clinical trials and marketing. But it is not expecting to hire any new staff in the near future. "It's a survival atmosphere," says

Christopher Price, chief executive of Nobex, a small biotech in Research Triangle Park, North Carolina.

"Demand will ebb and flow a little bit," says Primeau. "But there's still a strong demand for medicinal chemists."

Eugene Russo is a freelance science writer in Takoma Park, Maryland.



Povl Krogsgaard-Larsen



Christopher Price



Richard Scheller